Project title:	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space
Project number:	101061001
Project Acronym:	AD4GD
Type:	HORIZON-AG - HORIZON Action Grant Budget-Based
Work program topics addressed:	HORIZON-CL6-2021-GOVERNANCE-01

DELIVERABLE NO: D8.2
DATA MANAGEMENT PLAN

Due date of deliverable:	28/02/2023
Actual submission date:	06/03/2023
Version:	1.2
Main Authors:	Vasiliki Tsiompanidou, Claudia Martorelli, Adrian Quesada Rodriguez, Renata Radocz, Ana Maria Pacheco, Sébastien Ziegler, Stefan Schiffner

DOCUMENT METADATA

Project number	101061001
Project title	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space

Deliverable title	Data Management Plan
Deliverable number	D8.2
Deliverable version	1
Contractual date of delivery	28/02/2023
Actual date of delivery	06/03/2023
Document status	Final
Document version	1.2
Online access	
Dissemination	Public
Work package	WP8: Project Management and Coordination
Partner responsible	European Centre for Certification and Privacy (ECCP)
Author(s)	Vasiliki Tsiompanidou, Claudia Martorelli, Ana Maria Pacheco, Sébastien Ziegler, Stefan Schiffner
Editor(s)	Vasiliki Tsiompanidou
Reviewer(s)	Francesca Noardo
EC Project Officer	Lara Congiu

Abstract	The Data Management Plan provides guidance on data management and personal data protection principles that need to be followed in the course of the project. It explains how partners are designing their data management strategy and how they are planning to ensure compliance with legal and ethical requirements, while providing data in a FAIR manner. It also describes the approach with regards to intellectual property rights. The Data Management Plan is a living document that will be constantly reviewed and updated throughout the project's lifecycle in order to better reflect the partners' strategy at different points of the project.
Keywords	Data management, data protection, privacy, data security, FAIR data, intellectual property rights
Disclaimer	Views and opinions expressed in this deliverable are those of the author(s) only and do not necessarily reflect those of the European Union the United Kingdom or Switzerland. Neither the European Union nor United Kingdom nor Switzerland can be held responsible for them

DOCUMENT VERSION HISTORY

Version history			
Version	**Date**	**Modification reason**	**Modified by**
0.1	21/11/2022	Initial version of the document & ToC	Stea Miteva (ECCP)
0.2	06/02/2023	Initial draft	Vasiliki Tsiompanidou (ECCP)
0.3	10/02/2023	Expansion on legal requirements	Claudia Martorelli (ECCP)
1.0	28/02/2023	Review and final version submitted for review	Vasiliki Tsiompanidou (ECCP); Sébastien Ziegler (ECCP)
1.1	02/03/2023	Peer review	Francesca Noardo (OGC)
1.2	03/03/2023	Final version ready for submission	Vasiliki Tsiompanidou (ECCP); Sébastien Ziegler (ECCP)

ABBREVIATIONS

Abbreviation	Definition
AD4GD	AllData4GreenDeal, the current project
EU	European Union
EC	European Commission
IPR	Intellectual Property Rights
IoT	Internet of Things
AI	Artificial Intelligence
FAIR	Findable, Accessible, Interoperable and Re-usable
UN	United Nations
GEOSS	Global Earth Observation System of Systems
EOSC	European Open Science Cloud
API	Application Programming Interface
CitSci	Citizen Science
WP	Work Package
GDPR	General Data Protection Regulation
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessment
DPA	Data Protection Authority
DCA	Data Controllership Agreement
NIS	Network and Information Security Directive

DGA	Data Governance Act
WP29	Article 29 Working Party
EDPB	European Data Protection Board

Table of Contents

Table of Figures

1 EXECUTIVE SUMMARY

The Data Management Plan provides insight on AD4GD's approach with regards to data that will be generated and/or collected and/or processed within the context of the project. In particular, it analyzes the project's **dual approach**, as follows:

1. **Personal data**, focusing on developing solutions and policies that will ensure they remain adequately protected at all times, adopting a privacy by design and by default approach. The project has already taken into consideration the legal obligations prescribed to them by the GDPR and other relevant EU legislations and the present deliverable provides them with additional guidance on them. As the project progresses, a more in-depth analysis will ensue matching the partners to their specific obligations according to their role within the project and their envisioned action points.
2. **Non-personal data**, focusing on ensuring they are shared with the scientific community and industry in a FAIR manner. Interoperability forms a central notion of the project's objectives, which aim at creating a Green Deal Data Space that will be providing access to crucial datasets so that they can be re-used to meet the EU Strategy's goals related to the protection of the environment and the prevention of climate change.

In addition to the above, the present deliverable describes the project's **commitments towards the principles of Open Science and Open Data** to the greater extent possible, while providing additional guidelines as to the management of Intellectual Property Rights developed in the context of the project.

Annex I provides the **Questionnaire** that will be distributed within the Consortium in order to **further evolve the data management strategy of the partners, identify and address any upcoming needs**, as will be described in the following iteration of the Data Management Plan.

2 INTRODUCTION

2.1 OBJECTIVES OF THE PROJECT

In the past decade, environmental and sustainability concerns have become more and more of a priority for legislators and decision-makers all over the globe. In particular, as far as Europe is concerned, the European Commission has established a set of priorities to be reached until 2024, among which the European Green Deal[1]. The European Green Deal encompasses a number of goals related to the protection of the environment and increased sustainability, aiming at helping Europe become the first climate-neutral continent by 2050[2].

A major part of the above-described Green Deal is the establishment of a Green Deal Data Space that will assist the EU in harnessing the power of data towards a sustainable future. As such, AD4GD's main objective is to co-create and shape the European Green Deal Data Space as an open hub for FAIR data and standards-based services that support the key priorities of biodiversity, climate change, circular economy, deforestation, and pollution.

In order to achieve this, the project will focus on interoperability as a solution to the semantic and technology gaps, thus enabling multi-disciplinary and multi-scale access to data, processing services, and processing platforms. As such, the project's objectives have been defined as follows:

1. To **co-design a Green Deal Common Data Space** consisting of **interoperable** building blocks for heterogeneous data integration, artificial intelligence, Web APIs etc., using **semantic mapping** to allow for multiple existing data models and API standards to be fully integrated.
2. To ensure the **FAIR integration of CitSci** with other in-situ Earth observation data and INSPIRE data in the European Green Deal Data Space.
3. To enable heterogeneous IoT communication protocols and data format integration into a **common semantic model** for the climate-related, geospatial and environmental European Green Deal Data Space.
4. To **overcome data fragmentation** by combining Earth Observation data from satellites with other sources of data into a common climate-related, geospatial and environmental data space to support the European Green Deal Data Space.
5. To **enhance certainty, quality, and exploitability of heterogeneous data** by leveraging on data analytics, machine learning, and Artificial Intelligence.
6. To demonstrate through multi-scale, multi-criteria and multi-actor pilots the **applicability and added value of data fusion for improved accessibility decision-making** in the European Green Deal Data Space domains climate change, zero pollution and biodiversity.
7. To **research and demonstrate the potential of the AD4GD data space** concept from the core to the edge to increase the scalability, performance, and convergence of the use of high-performance computing, cloud, data, and artificial intelligence resources for Earth system modelling.

[1] European Commission, 'The European Commission's priorities; 6 Commission priorities for 2019-24' (16 July 2019), available at <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024_en> [accessed 19 February 2023].

[2] European Commission, 'A European Green Deal; Striving to be the first climate-neutral continent' (11 December 2019), available at < https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en> [accessed 19 February 2023].

8. **To upscale and sustain the AD4GD concept and a collaborative community** to support a highly scalable, comprehensive, and FAIR Green Deal European Data Space for citizens, researchers, policy, and decision makers.

2.2 PURPOSE OF THE DOCUMENT

The present document intends to shed light on the various categories of data that may be collected/processed or generated in the course of the project, whether falling under the category of personal data or not. In this context, it will focus on providing significant guidance as to the standards and principles that must be upheld during the project's lifecycle as well as after its completion. In particular, it will focus on the following aspects:

1. The **identification and description of the data** that may be generated and/or collected and/or processed in the course of AD4GD;
2. Guidance on how to ensure the project's data is **FAIR** (Findable, Accessible, Interoperable and Re-usable);
3. The analysis of relevant **Personal Data Protection and Privacy principles, legislation and guidelines**;
4. The analysis of **other ethical and legal provisions that are of relevance** to the project's goals and action plan, especially when they have the potential of affecting data sharing;
5. Guidance on how to approach **Intellectual Property Rights** that may result from the project, having in mind the Open Science and FAIR data principles.

2.3 METHODOLOGY

The present Data Management Plan explores the implications of the handling of data, personal or not, from a legal and ethical perspective. It is based on research on not only the GDPR, i.e. the main data protection instrument in the EU, but also on further applicable EU legislation, as is reported on Section 5 of the present deliverable. The regulatory framework includes not only existing but also anticipated legislation on a European level. In addition, it analyses the requirements for data to be FAIR and explains which is the project's initial approach towards it. Finally, it describes the Consortium's existing and upcoming strategy regarding the management of Intellectual Property Rights.

3 RELEVANT STANDARDS AND PRINCIPLES FOR THE DATA PROCESSING IN AD4GD

3.1 OVERVIEW

In line with the EU Strategy for Data, AD4GD's aspiration is to make a significant **contribution to the single market for data** to be created. As such, data lies at the heart of the AD4GD project. With the focus being on supporting establishment and operation of the European Green Data Space in particular, data can be deemed as the most crucial element with regards to AD4GD's course of action and its results, both during the project's lifecycle and beyond it.

In view of the above, the project is expected to not only collect and/or process data, but also to generate a number of datasets, as will be further analyzed in the next sections. AD4GD intends to approach all data processing and data generation performed within the context of the project in a way that is data protection and privacy-compliant, in alignment with additional ethical and legal requirements of relevance, including the Artificial Intelligence Act, as well as the FAIR data principles. Where applicable, AD4GD

intends to make available datasets of high quality in an accessible manner while guaranteeing interoperability.

In addition, as per 2021 UNESCO's Recommendation on Open Science[3], AD4GD aims at having its data meet the following conditions where possible:

a. **Available in a timely manner,**

b. **Through a user-friendly format,**

c. **Human and machine-readable,**

d. **Actionable,**

e. **In accordance with the principles of good data governance, stewardship, the FAIR principles,**

f. **Supported by regular curation and maintenance.**

Given the early stage of the project, the precise characteristics of the datasets that will be generated/collected/processed in the course of the project are yet to be determined, and will be reported in greater detail in the upcoming second iteration of the Data Management Plan. In order to assist partners, a relevant Questionnaire, as attached in Annex I of the present deliverable, will be shared with the Consortium. Using said Questionnaire, partners will be able to better reflect on, build and report on their data generation and processing activities, their own data management strategy, as well as their IPR-related exploitation ambitions.

3.1.1 RELEVANCE OF POTENTIAL DATA PROCESSING ACTIVITIES FOR THE PROJECT

As already explained, data plays a central role in the context of AD4GD. In line with one of the main goals of the project related to **semantic interoperability**, AD4GD is expected to generate datasets for tests and validation of data integration. In this way, it will be able to demonstrate the **applicability and added value of data fusion**, as well as the role it will hold with regards to enhanced accessibility in the decision-making process in the Green Deal Data Space.

Enabling the **combination and integration of data from heterogeneous sources in an interoperable, scalable and reliable manner**, AD4GD aims at rendering the datasets accessible to the knowledge centers, GEOSS portal, and other science services, as applicable. Aiming at identifying and addressing the barriers that limit the sharing of data and building blocks among institutions, AD4GD envisions to facilitate cross-organization data sharing, including between European institutions and international organizations such as WMO, GEO, and the UN system.

Taking the above into consideration, it becomes apparent that various data may be processed and/or generated within the project. Depending on whether data is personal or not, AD4GD shall adopt a different strategy, either focusing on ensuring privacy by design and by default or openness, availability, interoperability, accessibility and reusability accordingly.

[3] UNESCO, 'UNESCO Recommendation on Open Science' (2021) <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.

3.1.2 MAIN CATEGORIES OF POTENTIAL DATA

There are various categories of data that will be of relevance for the AD4GD project. As already explained, the exact categories of data that will be generated and/or processed within AD4GD will be reported in detail in the second iteration of the Data Management Plan. Nonetheless, a preliminary identification of the relevant data for AD4GD has been performed and has concluded on the following main categories of data:

1. **Personal data**: Personal data in the context of the project shall be mainly collected in the ways described below:
    a. **From contributors** to the datasets **and users** requiring access to the applications and services deployed in the project. As such, it will primarily involve names, usernames, email addresses and passwords while it may also require further identification information such as affiliated organizations and physical addresses. Such data will be processed with the utmost care, since the system envisioned will be designed with privacy concerns in mind while enabling dialogue and interoperability.
    b. **From citizens** as active sensors in a network of observations. Since geo-located data will lie at the center of the project, citizen's data may be incidentally collected from IoT devices. Of course, such data will be anonymized to the greater extent possible, while the project will also leverage on synergies thanks to the Big Data Value Partnership that will provide invaluable insights on privacy-preserving technologies and cybersecurity measures that can be implemented.

    Given the importance of personal data protection and privacy in the context of AD4GD, such concerns are taken into account as early as the **design phase so that strict data protection policies can be established**. As will be further analyzed in Section 5 of the present deliverable, any personal data that may be processed within the project will be handled considering the relevant data protection legislation and principles, and primarily data protection by design and by default, data minimization, the protection of data by anonymization and/or pseudonymization methods where applicable and transparency, especially when AI technology is utilized.

2. **Non-personal data**: AD4GD will be not only processing but also generating a number of datasets that do not involve personal data. The main categories of such data have been identified as follows:
    a. **IoT data, statistical data and community observations** that will be used to develop the algorithmic models to improve existing Earth observation models.
    b. **Governmental, commercial and volunteered data** that will be combined using adaptors and semantic mappings, which should themselves be accessible as FAIR services, aiming at improving the quality and reliability of the data.
    c. **Data coming from existing networks of in situ observatories** such as the ENVRI plus network, the European Citizen Science association and the INSPIRE community. Such data shall be used to design and implement an approach based on dialogue and co-creation so as to define the EV framework.
    d. **In-situ Observations, Socioeconomic Data and CitSci Data** will be integrated in order to design and cross-validate alternative models for generating environmental metrics.
    e. **Satellite and earth observations** that will be used to determine, following a multi-actor approach with the involvement of relevant stakeholders and users, a range of datasets that can support a thorough testing of the proposed semantic model and which can be effectively accessed for the project's case studies.

The above-described data is intended to be made **available** by the project, to the greater extent possible, in an **open and accessible manner**, in line with the Open Science and the FAIR Data principles.

4 FAIR DATA

AD4GD ambitions to not only support but also play a catalytical role in the design of the European Green Data Space. In order to achieve this, FAIR data is essential for the project's goals and aspirations.

Additionally, the project explicitly recognises the importance of Open Data and its FAIR handling and, thus, plans to **leverage on the EOSC infrastructure**, benefiting from its common pool for accessible, interoperable, reusable, and open research data. By joining its network, AD4GD shall commit to adhere to EOSC's policies for research and set principles related to the seamless access to data, FAIR data management, and reuse of data among others. In order to maximize impact, it also intends to **connect with EOSC-related projects and Horizon 2020 projects**, extracting a set of best practices and features that can be used to enable the interoperability of the various platform and the establishment of a FAIR data ecosystem.

The present section will provide an overview of the already identified points of actions to render the AD4GD-generated data FAIR, as well as the criteria those must meet[4]. Further analysis will follow in the upcoming second iteration of the Data Management Plan based on the extensive information that will be collected through the Questionnaire attached in Annex I which will be filled by all partners in the following months.

4.1 MAKING DATA FINDABLE

In order to make data and metadata findable, it is essential that it meets the following conditions[5]:

i. It is assigned a **globally unique and persistent identifier**;
ii. It is described with **rich standardized metadata;**
iii. It is registered or indexed in a **searchable resource**.

As such, AD4GD has already recognized the need for **standard agreements on metadata aspects for discovery, APIs and resource models** that will interact with the environmental service information system. It will aim to **incorporate already recognized vocabularies and standardized metadata**, while also establishing a **Metadata Working Group** that will be in charge of adequately formulating the metadata relevant for each dataset. Curating and tagging the data, linking as much as possible to existing standards (e.g. DCAT, GeoDCAT, ISO19115), will be highly useful so that the best-fitted data sources can be discovered and matched to the most appropriate stages of a scientific workflow.

4.2 MAKING DATA OPENLY ACCESSIBLE

Accordingly, once data is found, there needs to be the possibility to access it, whether without additional steps or after authentication and/or authorization has been concluded. In order for data to be considered accessible, it needs to meet the following criteria[6]:

[4] GO FAIR, 'FAIR Principles' available at: < https://www.go-fair.org/wp-content/uploads/2022/01/FAIRPrinciples_overview.pdf> [accessed 20 February 2023].
[5] Ibid no 4.

i. It is **retrievable by its identifier using a standardized communications protocol;**
ii. The protocol used is **open, free, and universally implementable;**
iii. The protocol used allows for an **authentication and authorisation procedure, where necessary**;
iv. It permits for **metadata to be accessible,** even when the data no longer is.

As already explained, the project aims at **benefiting from existing standardized metadata and communications protocols**. Leveraging on the knowledge obtained through similar Open Data Space initiatives and adopting a **co-design approach,** it intends to research, instantiate and test a multi-protocol proxy webservice accessible to relevant stakeholders and users, with the goal of expanding interoperability with Earth observation data.

What is more, AD4GD's goal is based on providing access to its findings and datasets to other relevant users and stakeholders co-developing the European Green Data Space and is, therefore, committing to propose solutions that will facilitate these procedures.

4.3 MAKING DATA INTEROPERABLE

Interoperability has been highlighted as the main component of the project's objectives. As such, they must meet the below described general conditions[7]:

i. Both data and metadata need to use a **formal, accessible, shared, and broadly applicable language** for knowledge representation;
ii. Both data and metadata need to use **vocabularies that are compliant with FAIR principles**;
iii. Both data and metadata need to include **qualified references to other data and metadata.**

In that sense, AD4GD links interoperability and reusability of FAIR data to clearly communicating its provenance and documenting the transformations undergone, as well as recording more traditional summaries of data quality to assess whether they are fit for new users.

In particular, the project commits to research, instantiate and test a **multi-protocol proxy webservice with the purpose of expending interoperability with Earth observation data,** while enabling and facilitating **interoperability among its APIs with IoT devices and other data sources,** such as openly available data. Within this action point, it will propose, develop and test **semantic interoperability approaches** suitable for incorporating IoT and other data streams currently not being harvested and available in contemporary data platforms. As a result, wider data spaces and new applications will be unlocked that will serve for data-driven mitigation and adaptation strategy as well as for policy design and monitoring.

Expanding on the above, AD4GD aims at enabling semantic interoperability, through the development of a common semantic model that will further develop the Essential Variables framework. In turn, the framework is intended to provide the reference vocabulary that will enable different components from different providers to interoperate and exchange data related to the European Green Deal. Said model will **implement semantic mappings with other standard and/or dominant models, which would enable the semantic integration of data** represented based on those models. It will be built on the basis of existing ontologies and vocabularies already available for the targeted domains, exploiting as much as possible existing INSPIRE, OGC and other standards, while extending them.

[6] Ibid no 4.
[77] Ibid no 4.

In order to achieve the above, the project will prepare a **semantic interoperability mapping** in order to discover **better connectivity solutions between the main data hosts**, facilitating the emergence of new applications that require multiple datasets from different sources. It will also seek advice from initiatives that are already active in the technical developments for the curation, citation and harmonisation of heterogeneous data, such as GBIF, in order to facilitate and expedite the interoperability of the solutions and datasets.

4.4 RE-USE OF DATA

Similarly, datasets generated within AD4GD need to be reusable, meeting the following minimum set of criteria[8]:

i. Both data and metadata need to be **richly described** with a plurality of accurate and relevant
ii. Attributes;
iii. Both data and metadata need to be **released with a clear and accessible data usage license**;
iv. Both data and metadata need to be **associated with detailed provenance**;
v. Both data and metadata need to meet **domain-relevant community standards**

The project's general approach that is based on **co-creation and co-design of the data space with the community of data providers and end-users** aims at a common semantic interoperability framework that will enhance the usability of data and their corresponding information so as to support of the European Grenn Deal strategies.

To enhance reusability, the semantic models that will be produced, defined as profiles that implement additional external models, need to be formally described. In such a case, specific constraints used by the semantic model on each re-used existing model can be automatically extracted and defined as profiles of each of these. Based on the formal definitions, future applications in related domains can reuse and declare interoperability with the resources of that semantic model. Within OGC, the development of a data model profiling tool is being planned, in order to allow such a formal description of standard data models profiles. It will support data requirements description, data validation and filling part of metadata. Such a profile will be standard itself and will represent an agreement about interoperability as well. It will be developed and tested for the AD4GD data management, supporting reusability of data and developed solutions.

Additionally, in the context of the project **a number of powerful re-usable tools will be developed, aiming at data harmonisation and retrospective annotation with quality-relevant information**. For this purpose, existing standards and ontologies will be exploited, wherever possible, including PROV, UncertML and QualityML, PPSR_CORE and the GEOSS GEOLabel. Standards and vocabularies for documenting provenance, QA and QC of remote sensing data, as well as for IoT sensors and networks (e.g. Sensor Things API, SensorML, the Semantic Sensor Network Ontology) will also be used.

Finally, AD4GD envisions a European Green Deal Data Space that will be defined in such a way that its components will be able to **serve the EC Knowledge Centres** when accessed or re-used by the Science Service. As such, the use cases defined in the project will demonstrate the **capability to create knowledge that can be transferred to the Knowledge Centres on Earth Observation, Biodiversity and Bio-economy, securing the long-term sustainability of the work**, benefiting from partners' already-existing relationships with such centers.

[8] Ibid no 4.

5 GEOSS PRINCIPLES OF DATA MANAGEMENT AND DATA SHARING

In order to best align the project's activities with the Open Science principle, AD4GD will also consider the complementary guiding principles provided by the Group on Earth Observations with regards to data management[9] and data sharing[10]. As such, as far as **data management** is concerned, the Consortium will design its activities bearing in mind the following:

1. **Discoverability,** rendering the data findable, as described in Section 4.1;
2. **Accessibility,** as described in Section 4.2;
3. **Usability,** expanding on the re-usability principle described in Section 4.4;
4. **Preservation,** requiring that data remains protected from loss and is preserved for future use, in accordance with a preservation plan that will also ensure that integrity, authenticity and readability is verified for both data and metadata;
5. **Curation**, ensuring that corrections, updates and reprocessing are possible, while enabling citations of data based on persistent and resolvable identifiers.

Similarly, the GEO has formulated the following **data sharing principles**, promoting:

1. **Full and open exchange** of data, metadata and products shared withing GEOSS;
2. Availability of all shared data, metadata and products **with minimum time delay and at minimum cost**;
3. Availability of all shared data, metadata and products **free of charge or at a cost no more than the cost of reproduction for research and education purposes.**

AD4GD will take into account the above-described principles and shall incorporate them as much as possible to its activities and architecture, while respecting legal and ethical requirements, as analyzed in Section 6 and potentially updated in future iterations of the Data Management Plan in accordance with the evolution of the regulatory and ethical framework. The above principles will also be balanced against the partners' potential Intellectual property rights as will be described in Section 7 of the present.

6 ETHICAL AND LEGAL ASPECTS

6.1 TASK MANAGEMENT WITHIN THE PROJECT

The main technical and implementation components of the project have been divided into individual work packages and allocated to task leaders, according to their intended purpose and expected outputs. The figure below provides a high-level overview of the distribution of responsibilities and tasks among the different work packages.

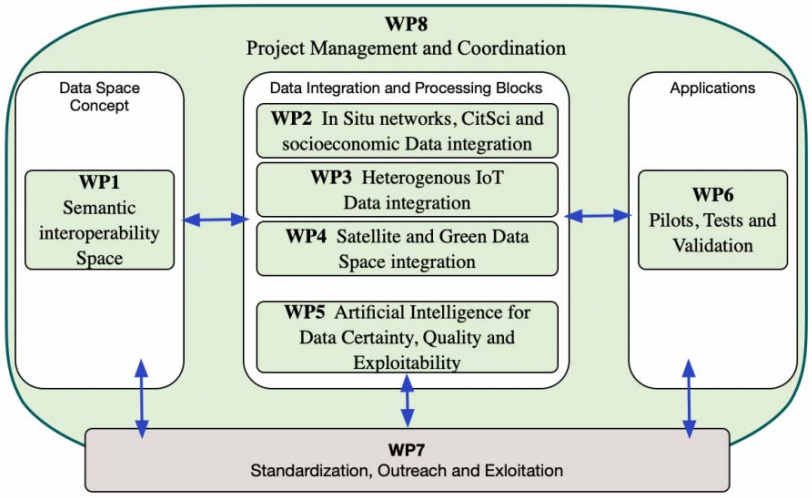

Figure 1 Workplan structure and interactions

[9] GEO Data Manag l5).
[10] Data Sharing W les post 2015' (10 March 2014).

Based on the above, WP1 shall provide a stakeholder-driven method to manage data in the context of the project, which will allow for a more detailed dataflow mapping and allocation of responsibilities. As the work on general data management plans progresses, the data mapping will be continuously updated over the course of the project.

For the **identification of the roles of the partners with regards to the data processing activities,** as well as the more precise definition of said data processing activities, partners will be provided with **Data Management Questionnaires (Annex I)**. Annex II also provides explanations as to what is meant by the main terminology used, including but not limited to the concepts of "personal data" and "data processing".

By the time the next version of the DMP is due, and on the basis of the Questionnaires distributed, as the partners shall provide the relevant information, this section will be updated and completed accordingly to reflect the true state of the project's activities.

6.1.1 DATA PROTECTION OFFICER

Articles 37-39 GDPR define the role of the Data Protection Officer, a figure distinguished for their expert knowledge of data protection law and practices. The DPO is, generally, in charge of monitoring the application of the GDPR within an organization, while informing and advising controllers on how to comply with obligations stemming from the applicable data protection provisions.

As such, and to ensure compliance with the relevant data protection legislation, AD4GD, in the context of Task 8.3, has appointed a **DPO for the project**, who will be in charge of:

- **Coordinating with the respective DPOs of the data controllers** involved in the project, as will be identified through the Questionnaires displayed in Annex I;
- **Overviewing the project's compliance** with the GDPR and other current and upcoming regulatory obligations;
- **Periodically reporting on the project's compliance** with applicable data protection norms in the course of project development;
- **Performing a formal DPIA** that will analyze any risks posed for personal data, that will be included accordingly in the next iteration of the Data Management Plan;
- **Providing recommendations,** where applicable, to partners regarding their data management strategy within the project.

6.2 RESEARCH PARTNERS AS DATA CONTROLLERS AND/ OR DATA PROCESSORS

This section provides additional clarifications regarding the potential role of the partners within the project, from a GDPR perspective in the context of the research, namely as **data controllers, joint controllers or data processors/sub-processors.** Depending on the activities carried out in the research project, each partner can assume multiple roles and responsibilities from a regulatory perspective.

The understanding of the terminology and of the provisions of the GDPR is important in order to assure alignment in the communication within the consortium, but also to be able to correctly distinguish personal from non-personal data. In order to facilitate this, **an overview of relevant definitions and concepts of the European Data Protection legal framework can be found in Annex II** of this deliverable.

6.2.1 ROLES AND RESPONSIBILITIES IN THE PROJECT

As previously stated, each research partner can assume multiple roles and responsibilities according to the activities carried out in the project. For this reason, partners are required to define and communicate **how their work is organized, what personal data are/will be collected and processed to complete the relevant dataflow mapping, as well as their role in the project, so as to identify the controller(s) and processor(s)**. The Questionnaires provided to the partners will assist precisely in clarifying their role in personal data processing activities carried out in the context of the project and will be reported accordingly.

The AD4GD project **envisages the integration of data from different sources and the extraction of knowledge using AI** to facilitate decision-making in the context of environmental issues. WP2, 3 and 4 are dedicated to the integration of data collected from *in situ* networks and CitSci, socioeconomic, IoT and satellite data as well as data generated from third party services. In these phases research partners will identify which data need to be integrated to tackle challenges in the Green Deal priority areas. WP5 is dedicated to the development and use of AI to ensure data quality and extract knowledge from the integrated data space. To the extent personal data are processed in such operations, research partners will need to abide by the relevant data protection legislation.

The AD4GD project also includes **three pilots to demonstrate the value that the data integration approach thus developed can add in tackling urban and rural sustainability challenges.** The pilots will deal with water pollution monitoring, monitoring and optimization of biodiversity corridors and the enhancement of greenhouse gas emissions monitoring through dynamic calibration of low-cost sensors. Where personal data is involved, the actors participating in the pilots will need to follow the guidelines analyzed below.

As already explained, the **DPO of the project will coordinate the data protection policy** at the project's level and **overview its compliance with the relevant regulatory obligations, both current and upcoming**. As such, the DPO can support, bring solutions and guidance at consortium level on how to ensure compliance with the relevant legislation. It can also facilitate horizontal cooperation and the sharing of good practices between the partners and their respective DPOs. Data sharing and other associated tasks will also be coordinated as part of these activities.

6.2.2 INITIAL INSTRUCTIONS AND OBLIGATIONS TO BE RESPECTED

In the course of the project, partners may act in the capacity of data controllers, whether jointly or not, as well as in the capacity of data processors. The present section shall provide guidance regarding the main obligations when assuming either of these roles.

6.2.2.1 AS CONTROLLERS

Entities acting as controllers bear the highest level of responsibility with regard to compliance with data protection provisions. A controller has to abide by the following obligations:

- **Purpose limitation**: Personal data shall be accessed only for the purpose of the project and in line with the project's associated agreements (Grant Agreement, Consortium Agreement). No further processing of personal data by the partners for their own purposes or for purposes of third parties is permitted unless otherwise stated in data protection legislation;
- **Data minimization**: All personal data introduced, transferred, or processed within the project must be adequate in relation to the purposes of the project. No personal data shall be processed or transferred unless necessary;

- **Storage limitation**: Personal data shall only be retained as long as this is absolutely necessary for the performance of the project's objectives and purposes. All personal data should be pseudonymized as soon as possible and anonymized, if possible, to carry out the project's objectives;
- **Accountability**: Controllers shall ensure and document that the activities carried out comply with applicable data protection laws and implement any necessary technical and organizational measures;
- **Data security**: Controllers shall ensure data security, integrity and confidentiality, avoiding to the greater extent possible accidental or unlawful destruction or loss, alteration, unauthorized disclosure of or access to data;
- **Data protection by design and by default**: Controllers shall consider compliance with data protection regulations from the early stages of processing, so that the processing is designed in a way that only the minimum necessary amount of data is processed;
- **Relations with processors**: Controllers shall appoint only processors that can ensure compliance with the GDPR, **through a written contract**, laying down clear rules, limitations and obligations for the processing activity;
- **Duty of cooperation**: Controllers shall cooperate, when requested, with the Data Protection Authority;
- **Data breaches notification**: Controllers shall notify data breaches to the competent DPA and to data subjects, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons.

6.2.2.2 AS JOINT CONTROLLERS

Whenever two or more partner organizations jointly determine the purposes and means of the processing in the context of the AD4GD project, they shall:

- **Share the responsibilities** mentioned in the previous section;
- **Enter into a Data Controllership Agreement**, in which they shall in a transparent manner determine their respective responsibilities for compliance with the GDPR and reflect their respective roles and relationships vis-à-vis the data subjects. The essence of the Data Controllership Agreement shall be made available to data subjects.

6.2.2.3 AS PROCESSORS

Partners performing data processing on behalf of a controller will act as processors and shall act only under the control and documented instructions of the controller. Furthermore, they shall abide by the following obligations:

- **Confidentiality**: Ensuring that persons authorized to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality;
- **Sub-processors**: Engaging another processor only with prior written authorization of the controller. A sub-processor can be appointed using a written contract which places on the sub-processor the same data protection obligations placed on the processor by the controller;
- **Deletion or return of data**: At the choice of the controller, when the relationship ends, deleting or returning to the controller all the personal data relating to processing and deleting existing copies unless Union or Member State law requires storage of the personal data;
- **Security measures**: Adopting all the technical and organizational measures to ensure a level of security appropriate to the risk of processing;

- **Record of processing activities**: Maintaining a record of any processing activities carried out on behalf of the controller;
- **Duty of assistance**: Assisting the controller in ensuring compliance with the controller's obligations;
- **Duty of cooperation**: Cooperating, when requested, with the DPA.

6.3 DATA PROTECTION FUNDAMENTALS

As previously explained, the overall objective of the project is to co-create and shape the European Green Deal Data Space to deliver open access to data to address climate change and environmental issues. This can be achieved through the combination and integration of environmental information gathered from heterogeneous sources and by making it accessible to a variety of actors. In practice, the project will use data that may be publicly available in the form of open data, but will mainly rely on geo-located data, integrating CitSci contributions.

As such, it has already been identified from the start of the project that potential risks of tracking and identifying individuals should be considered. Personal data such as usernames, email addresses and passwords to access the specific applications and services deployed in the project need to be taken into account as well. As a result, whenever personal data are involved, the project will follow strict personal data protection policies, in line with the GDPR and other norms, as applicable. The present section provides an overview of the main principles involved in the management of data within the project for the partners' consideration when designing their data management strategy.

6.3.1 PERSONAL DATA PROTECTION PRINCIPLES

With regards to the personal data protection principles, the project has considered not only the relevant legislation, but also Guidelines published by the EDPB. In addition to the obligations specified above for data controllers and data processors, below is a number of general principles to be respected by all partners:

1. **Lawfulness**: A legal basis must have been identified for the personal data processing, in accordance with Art. 6 (1) GDPR, namely:
    a. The data subjects' specific, freely given and explicit consent;
    b. Necessity for the performance of a contract;
    c. Necessity for the compliance with a legal obligation;
    d. Necessity for the performance of a task carried out in the public interest;
    e. Necessity for the purposes of the legitimate interests pursued by the controller or others, except if overridden by the data subjects' fundamental rights and interests;
    f. Necessity to protect the data subjects' or other natural persons' vital interests.
2. **Fairness**: Personal data must be processed in a manner that fairly balances the data subjects' rights and freedoms against the controllers' interests.
3. **Transparency**[11]: Personal data must be processed in a transparent manner, providing data subjects with all necessary information with regards to their personal data, the processing activities, the purposes of the processing, as well as the parties with which they are shared.
4. **Accountability**: Compliance with the relevant personal data protection principles and legal provisions must be demonstrated.

[11] WP29, 'Guidelines on transparency under Regulation 2016/679' (WP260 rev.01, 11 April 2018) [26].

5. **Purpose limitation**: Data must be processed only for specified, explicit and legitimate purposes. Any further processing must be compatible with the original purposes, among which scientific research purposes, also in accordance with national legislations.
6. **Data minimization and storage limitation:** Only the data that are strictly relevant and necessary to the purpose of the processed may be processed and they must not be retained than what is required to attain said purpose.
7. **Data protection by design and by default:** Data protection must be considered from the beginning of the project and adequate, state-of-the art measures must be adopted, to the greater extent possible, in order to protect data subjects' rights and freedoms[12].
8. **Accuracy, integrity and confidentiality of the data**: Personal data processed must remain accurate, up-to-date, reflecting their true status. It must be processed in a secure manner, adopting the technical and organizational measures that will help protect the data against unauthorised or unlawful processing and against accidental loss, destruction or damage.

Based on the above principles, partners are recommended to consider and adopt, where applicable, the following measures and practices:

1. **To identify their role within the project** as a data processor and/or data controller, the data processing activities they will be carrying out, the purposes and the legal basis for each of them;
2. **To minimize the use of personal data** as much as possible, prioritizing the use of anonymized or aggregated data;
3. **To adopt adequate technical and organizational measures,** including anonymization and/or pseudonymization of personal data where applicable;
4. **To accurately record information they collect or receive** and adopt procedures to ensure that inaccurate personal data can be easily erased or rectified;
5. **To adopt a data protection by design and by default strategy.** Within this context, partners must, in particular, consider the traceability of any algorithms for data validation, strict access control procedures to data and an overall security by design approach**;**
6. **To conduct a DPIA,** where applicable, according to the guidelines provided in Section 5.4 of the present deliverable;
7. **To certify their data processing activities** in accordance with a recognized certification scheme, where applicable.

5.3.2. DATA SUBJECT RIGHTS

Art. 8 of the European Charter of Fundamental Rights[13] provides the core elements of the right to personal data protection, which are further developed by the GDPR, which in turn establishes new rights for data subjects. As a result, the partners involved in the AD4GD project must commit to respect, guarantee and facilitate the exercise of those rights, where applicable, as follows:

a. **Right to information**, permitting them to learn, among others, who is collecting and processing their data, for which purpose and on which legal grounds, the duration it will be kept and with whom it shall be shared, unless otherwise stated in the legislation;

[12] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (version 2.0, 20 October 2020) [35].
[13] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.

b. **Right to access their personal data and obtain a copy** of the information referring to them[14];
c. **Right to rectification** of their personal data where it is inaccurate or incomplete;
d. **Right to object** to the processing of their data where the data was not collected directly from them, unless compelling legitimate grounds override their interests and rights;
e. **Right to erasure ("right to be forgotten"),** giving them the right to erase their data, unless an exception applies;
f. **Right to restrict** the processing of their data;
g. **Right to data portability,** where applicable, in a commonly used and machine-readable format[15];
h. **Right to not be subject to automated decision-making and profiling,** unless it is necessary for entering into, or performance of, a contract between the data subject and a data controller or it is authorised by Union or Member State law to which the controller is subject or it is based on the data subject's explicit consent. In the latter case, data subjects must be at least able to obtain human intervention.

5.4. DATA PROTECTION IMPACT ASSESSMENT

A DPIA is the process designed to describe the data-processing performed, assess its necessity and proportionality, and help identify and manage any risks to the rights and freedoms of natural persons.

The DPIA is required when the processing is "likely to result in a high risk to the rights and freedoms of natural persons". The rights and freedoms of data subjects primarily concerns the rights to data protection and privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience, and religion.

In order to facilitate understanding when a DPIA might be required, the Working Party 29's guidelines provide a common European Union list of **processing operations for which a DPIA is mandatory**, namely[16]:

- **When using evaluation or scoring methods, including profiling**, and predicting, in particular involving information related to the data subject's performance at work, economic situation, health, personal preferences or interests, behaviour, location or movements.
- **When employing automated decision-making** with legal or similar significant effect concerning the natural person.
- **When systematic monitoring of data subjects is performed**, including processing of data collected through networks or from a publicly accessible area.
- **When processing sensitive data or data of a highly personal nature,** including special categories of personal data and data relating to criminal convictions, as defined in Art. 9 and 10 of the GDPR respectively.

[14] EDPB, 'Guidelines 01/2022on data subject rights - Right of access' (version 1.0, 18 January 2022) [46].
[15] WP29, 'Guidelines on the right to data portability' (WP 242 rev.01, 5 April 2017) 9-10.
[16] WP29, 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679' (WP248 rev.01, 4 October 2017) 9-11.

- **When processing data on a large scale**, in particular taking into account the number of data subjects concerned, the volume of data processed, the duration and the geographical extent of the processing.
- **When matching or combining datasets**, originating from two or more data processing operations which were performed for different purposes and/or by different data controllers, exceeding the reasonable expectations of data subjects.
- **When processing data concerning vulnerable categories of data subjects**, such as minors of age.
- **When using in an innovative manner or when applying new technological or organisational solutions**, such as AI or IoT.
- **When the processing itself prevents data subjects from exercising a right or using a service or a contract.**

Based on the above, it needs to be highlighted that the DPIA s**hould be carried out before the data processing is performed,** considering the possible implication from the beginning even if the processing operation have not been clearly defined yet. In the latter case, the DPIA will need to be **updated** once the data processing officially starts and maintained **throughout the lifecycle of the project,** in order to reflect the actual state of the activities involved and maintain compliance. The DPIA must at least include a description of the envisaged processing operations and purposes, an assessment of compliance with the principles of necessity and proportionality, as well as the risks to the rights and freedoms of data subjects including the measures to address them.

According to the AD4GD research proposal, the project's DPO has taken on the performance of a formal DPIA as part of their duties with regards to the project. Once the partners' data processing activities that may involve personal data are identified, **a DPIA will be performed**, as will be reported in the second iteration of the Data Management Plan. Said DPIA will be **maintained and updated throughout the project's lifecycle** in order to ensure that the data processing activities performed remain **compliant**.

5.5. ETHICAL PRINCIPLES AND REGULATORY FRAMEWORK

As already attested, AD4GD and the activities related to the project shall comply with ethical and legal principles, standards and regulation. This includes undertaking activities in compliance with a number of ethical principles, of which the main ones can be found below:

a. **No data collected will be sold or used for any purposes other than the AD4GD project**.
b. **Any additional personal data obtained, but not intentionally collected, during the course of the pilots will be immediately erased.** Only relevant personal data will be collected and processed, and it will be anonymized as soon as possible given the project's goals and aspirations.
c. **No personal data will be disclosed or otherwise made available** beyond the project's objectives and the purposes for which it was collected. Any data sharing within AD4GD is subject to legal requirements and data sharing agreements will be signed to clearly define the roles, rights and obligations of each partner involved.
d. **Where natural persons are to be recruited as participants to the activities of the projects (such as surveys), appropriate measures will be adopted to ensure their privacy is respected and their information remains confidential**. Similarly, an adequate framework must be implemented to ensure no discrimination takes place in the context of the project.

e. **Where publications or other dissemination activities are to take place, no personal data must be shared or disclosed**. Anonymization techniques and the use of aggregated data will be of utmost importance at this stage.

5.5.1. ETHICS GUIDELINES FOR TRUSTWORTHY AI

As already stated, the European Commission has recognized that the use of AI can have a significant impact on the achievement of the goals of the Green Deal[17]. The AD4GD project will be deploying AI to assess the consistency and reliability of data sources in the European Green Deal Data Space, and to extract knowledge so to allow better informed decision-making to deal with a number of key environmental issues. In such cases, additional ethical guidelines must be upheld, in addition to those described above, in order to ensure adequate compliance.

As such, in the context of AD4GD, AI will be developed in a way that is respectful of EU rules and values. In particular, the Ethics Guidelines for Trustworthy AI[18] provide a framework for the development and use of AI in respect of the Charter of Fundamental Rights of the European Union and other relevant international human rights law. Research partners shall develop and use AI respecting the requirements set, as summarized below:

- **Human agency and oversight:** Human oversight shall ensure that AI fosters fundamental rights and supports user autonomy in decision making. Generally, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.
- **Technical robustness and safety:** Risk prevention and mitigation shall be considered in the development of AI so to ensure that AI is secured from malicious attacks and behaves as intended, minimizing unexpected harm. The results of AI systems shall be accurate, reproducible and reliable.
- **Privacy and data governance:** Adequate data governance shall be adopted so to ensure that AI systems guarantee privacy and data protection throughout a system's entire lifecycle. The quality and the integrity of processes and datasets used to train AI systems shall be verified and documented at each step such as planning, training, testing and deployment.
- **Transparency:** The elements that lead to AI system's decision (e.g. the datasets, the processes, the algorithms used) shall be documented to allow traceability and increase transparency. The technical processes of an AI system as well as related human decisions (e.g. application areas of a system) shall be explainable.
- **Diversity, non-discrimination and fairness:** Datasets used by AI systems shall be free from any identifiable and discriminatory biases. Oversight mechanisms shall be adopted to prevent that the development phase (e.g. algorithms' programming) may suffer from unfair bias. AI systems shall be user-centric, so to achieve high accessibility standards.
- **Societal and environmental well-being:** AI systems should be used to benefit all human beings, including future generations. The entire system's lifecycle should be as environmental friendly as possible, choosing options that involve less resource consumption. The impact of the system on institutions, democracy and society at large shall be considered carefully.

[17] European Commission, 'On Artificial Intelligence - A European approach to excellence and trust' (White Paper) COM (2020) 65 final, 2.

[18] High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (8 April 2019) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419> accessed 14 February 2023.

- **Accountability:** Mechanism shall be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use.

5.6. SUSTAINABLE DATA GOVERNANCE FOR THE GREEN DEAL

The AD4GD project aims at co-creating and shaping the European Green Deal Data Space to support key priorities of biodiversity, climate change, circular economy, deforestation, and pollution. The project's data management shall be based on sustainable data governance, whose objective is to benefit the European economy and society as a whole. Therefore, the AD4GD project commits to the objectives of the European Data Strategy and to the respect of relevant current and upcoming regulations and policy papers.

The scope of this section is to identify the main legislative provisions and requirements that the project shall abide by so to achieve a sustainable data governance for Green Deal.

5.6.2. EPRIVACY DIRECTIVE AND REGULATION

The ePrivacy Directive[19] was initially put into force to regulate the use of cookies in websites and provide an initial view of the data protection and privacy provisions in the sector of electronic communications. However, the Directive failed to meet its goals, hence the GDPR was put into force followed by the upcoming ePrivacy Regulation[20].

With regards to the data processing that may be performed within the context of AD4GD, the following provisions of the upcoming Regulation are of the most relevance:

- **Article 6,** providing the conditions of processing, differentiating among the following cases:
  - **When processing electronic communications:**
    - The processing must be necessary to achieve the transmission of the communication, as long as the criteria of necessity and proportionality as to the retainment period are met, or
    - The processing must be necessary to maintain or restore the security of the network or service or to fix technical errors.
  - **When processing metadata:**
    - The processing must be necessary to meet mandatory quality of services, or
    - The processing must be necessary for billing and interconnection payments, for detecting or ceasing fraudulent or abusive actions, or
    - The processing must be based on the users' consent for the already specified purposes
  - **When processing the content of electronic communications:**
    - Processing must be conducted for the sole purpose of providing specific services to end-users, as long as they have provided consent, recognising that said processing is indispensable, or
    - Processing must be necessary for the specified purposes for which the users have provided consent and the Supervisory Authority has authorised it.

[19] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L 201/37.

[20] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' COM(2017) 10 final.

- The requirement to **erase or anonymize the communications content and metadata** once the purposes have been concluded.
- **Article 8,** setting out a strict framework of conditions under which the processing and storage of information from end-users' equipment is allowed.
- **Chapter III,** dedicated to the rights of end-users to control the sending and reception of electronic communications to protect their privacy, guaranteeing anonymity[21] and its limitations[22], while providing the conditions under which end-users may be included in publicly available directories.

5.6.3. DATA GOVERNANCE ACT

The Data Governance Act[23], applicable as of 24 September 2023, is **part of the European Strategy for Data** envisaged by the European Commission, a central pillar of the European Green Data Space as well. It is aimed at **supporting the creation and development of European data spaces** leveraging on the collaboration between private and public players, in sectors such as health, environment, energy, agriculture, mobility, finance, manufacturing, public administration and skills.

Given the project's envisioned role in the establishment of the European Green Deal Data Space and the **DGA's objective of increasing data availability, building trust in data sharing and overcoming technical obstacles** to the reuse of data, **the DGA is of high relevance to AD4GD.**

In particular, based on the DGA, a single set of provisions are defined regulating the **conditions for the re-use of data held by public sector bodies which are protected on grounds of:**

(a) commercial confidentiality;

(b) statistical confidentiality;

(c) intellectual property rights of third parties; or

(d) the protection of personal data, insofar as such data fall outside the scope of Directive (EU) 2019/1024.

The **re-use of such data is encouraged** through the establishment of transparency obligations upon public bodies, i.e. the State, regional or local authorities, bodies governed by public law or associations formed by one or more such authorities, or one or more such bodies governed by public law, and of simplified procedures which can be followed by subjects interested in accessing them (e.g. the establishment of single information points). **Agreements pertaining to the re-use of such data which grant exclusive rights or restrict the availability of data for re-use by other entities are prohibited, unless specific conditions are met.**

5.6.4. DATA ACT

The Proposal for the Data Act[24] also forms part of the European Strategy for Data, complementing the dispositions of the DGA. While the DGA creates the processes and structures to facilitate data sharing and

[21] Article 12 of the draft ePrivacy Regulation.

[22] Article 13 of the draft ePrivacy Regulation.

[23] Regulation (EU) 2022/868 of The European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) [2022] OJ L 152/1.

[24] European Commission, 'Proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' COM (2022) 68 final. The proposal is undergoing the ordinary legislative procedure

availability, **the Data Act defines the actors who can create value from data and under which conditions,** in particular with regard to the data generated by IoT devices. In addition, the Data Act provides **specific guidance on interoperability requirements** to be complied with by operators of data spaces. In the context of the AD4GD project, this legislative proposal acquires relevance with reference to the environment-related data collected through IoT sensors as well as with regard to the provision of the envisioned integrated European Green Deal Data Space.

The Data Act introduces the **principle of data accessibility by default, requiring that IoT products are designed so to ensure that data generated by their use will be easily accessible** by default. Furthermore, the user of such devices can request from the data holder to make available the data to third parties, who shall ensure that the data are accurate, complete, reliable, relevant and up-to-date.

The Data Act also introduces **essential requirements regarding interoperability** to be complied with by operators of data spaces (Art. 28), namely that:

- the **dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty shall be sufficiently described to allow the recipient to find, access and use the data**;
- the **data structures, data formats, vocabularies, classification schemes, taxonomies and code lists shall be described in a publicly available and consistent manner**;
- the **technical means to access the data, such as application programming interfaces, and their terms of use and quality of service shall be sufficiently described to enable automatic access and transmission of data between parties, including continuously or in real-time in a machine-readable format.**

In this regard, the Commission is expected to adopt additional guidelines laying down interoperability specifications for the functioning of common European data spaces. In such case, AD4GD will consider them when designing its interoperability framework.

5.6.5. DATABASE DIRECTIVE

The Database Directive[25] is crucial to the AD4GD project as it **encourages the enrichment and fusion of existing datasets** which are currently created and managed within different silos. Since the AD4GD project is expected to generate datasets for tests and validation of data integration and its semantic interoperability space, such provisions will significantly facilitate it.

In particular, the Database Directive **protects both analogue and digital databases by copyright** if they are original because of the way their content is selected or arranged. In that sense, database is defined as "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means"[26]. Non-original databases (e.g. databases of scientific publications or of laws) can also be protected if the investment in obtaining, verifying and presenting the data was substantial. The protection of non-original databases (the *sui generis* right) takes the form of a property right which allows the maker of the database to prevent extraction and/or re-utilization of the

and it must be approved by the European Parliament and the Council of the EU. Once adopted, the Data Act will be directly applicable in all EU Member States one year following the publication in the Official Journal of the European Union.

[25] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 077/20.

[26] Art. 1 (2) of the Database Directive.

whole or of a substantial part of the contents of that database. The copyright and the *sui generis* right may both apply if the conditions of protection for each right are fulfilled.

The rights recognised to the author and/or the maker of the database may be limited in specific cases, such as:

- in the case of reproduction for private purposes of a non-electronic database;
- where there is use for the sole purpose of illustration for teaching or scientific research;
- where there is use for the purposes of public security of for the purposes of an administrative or judicial procedure.

However, it is clarified that **the *sui generis* database right does not apply to databases resulting from data generated or obtained by IoT devices.**

5.6.6. OPEN DATA DIRECTIVE

To create an integrated data space, the AD4GD project will use datasets provided in the form of open data, as it plans to benefit from the EOSC infrastructure and its pool of data. Furthermore, one of the project's main objectives is to share its findings and datasets to encourage the creation of the European Green Data Space. Against this backdrop, the **Open Data Directive**[27] **is particularly relevant in shaping the data management of the project.**

The Open Data Directive sets out **minimum rules on the re-use of data held by the public sector and of publicly funded research data made freely accessible through repositories.** The European Union data strategy aims at **unlocking the potential of open data** (e.g. data presented in open formats that one can use freely and share for any purpose), which **should be made available in formats that are open, machine readable, accessible, findable and reusable, complete with their metadata.**

In particular, **publicly funded research data can be reused for commercial or non-commercial purposes in cases where they are already made publicly available via an institutional or subject-based repository**. Based on the Open Data Directive, EU countries must adopt policies and take action to make publicly funded research data openly available, following the principle of **'open by default' and support the dissemination of research data, in accordance with the FAIR data principle.**

Concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests must be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'.

As the AD4GD project is aimed at producing integrated datasets on earth observation and environment data, it is worth mentioning that **the European Commission is in the process of adopting a list of high-value datasets**[28] **which should be made available in machine-readable formats and free of charge through APIs,** having started the relevant consultations in May 2022**.** The datasets will be selected from within 6 thematic categories, which include datasets on earth observation and environment. AD4GD shall closely monitor the evolution of said list and, where required, will adapt its activities accordingly.

[27] European Parliament and Council of the European Union, "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information," June 20, 2019, http://data.europa.eu/eli/dir/2019/1024/oj/eng, [Last accessed 27 February 2023].

[28] The list is still in preparation, updates can be found at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12111-Open-data-availability-of-public-datasets_en.

5.6.7. AI ACT

The AD4GD project plans to develop and use AI to ensure data quality and to extract usable knowledge from collected data. The dispositions of the Proposal for an AI Act[29] should be considered in the development of such technology as it sets a single set of rules **governing the development, placement on the market and use of AI systems[30] in the Union on the basis of a risk-based approach.**

The proposal provides a **risk-based approach to define "high-risk" AI systems**, which can give rise to significant risks to the health and safety or fundamental rights of persons, and it bans particularly harmful AI practices as contravening EU values. The **risk management approach** provided by the AI Act is aimed at testing, identifying and analyzing any foreseeable risks, evaluate them and adopt mitigation measures. In that context, data governance and data management practices are an essential part of the proposal, focusing on data that is relevant, representative, free of errors and complete.

At the same time, the need for **traceability, transparency and interpretation of outputs** has resulted in the requirement that for high-risk AI systems adequate technical documentation must be drafted, maintained and updated frequently. Of course, AI systems **must always permit human oversight**, achieving an appropriate level of **accuracy, robustness and cybersecurity** throughout their lifecycle.

Furthermore, the European Commission is expected to further clarify and adapt a list of approaches and techniques for the development of AI in Annex I of the proposal in line with new technological developments. Any updates to the AI Act will be closely monitored in order to ensure the project's compliance.

5.6.8. REGULATION ON THE FREE FLOW OF NON-PERSONAL DATA

As clarified in the previous sections of this deliverable, non-personal data are going to be processed in the context of the AD4GD project. The provisions set by the Regulation on the free flow of non-personal data[31] should therefore be considered to ensure that the processing is in line with these requirements.

The Regulation has the objective of **fostering the free movement of non-personal data across EU countries and IT systems in Europe.** The Regulation allows Member States to adopt localization requirements, including any legal or administrative measure which requires the processing of data to take place in a specific EU territory, only on the basis of public security grounds.

With the aim of ensuring enhanced transparency, the Regulation requires Member State to establish a **national online single information point containing all up-to-date localization requirements** and to appoint a single contact point to facilitate the cooperation with counterparts in other EU countries.

5.6.9. NIS DIRECTIVES

The first European Network and Information Security Directive on 'measures for a high common level of security of network and information systems across the Union' ('NIS1'),[32] from 2016, was the first piece of

[29] European Commission, 'Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts' COM(2021) 206 final.
[30] Art. 3(1)(1) defines "AI system" as a "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with."
[31] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L 303/59.

EU-wide legislation on cybersecurity aiming at achieving a high common level of cybersecurity across the Member States. It focused on three main pillars, namely the **enhancement of national cybersecurity capabilities, cross-border collaboration and the supervision of critical sectors providing public services.**

The rapidly changing technological framework quickly demonstrated the need for an updated version that would provide for modernized standards in all sectors. As such, the second directive 'measures for a high common level of cybersecurity across the Union', from 2022, **NIS2**[33] **extends the scope of application to include not only public and private entities deemed essential, but also important entities,** excluding only micro and small enterprises, unless they are offering public order services, under certain conditions.

The contribution of AD4GD to the Green Deal Data Space may result in its consideration under the following definition. As such, the project already considers the obligation to **ensure that cybersecurity tools and measures adopted sustain the general availability and integrity of the public core of the internet and is designing its solutions accordingly.**

6. PUBLICATIONS AND IPR GUIDELINES

AD4GD will generate a range of data and knowledge, some of which will be confidential and some of which will be for dissemination and communication to the public. IPR, within this context, must be deemed as meaning any IPR including, but not limited to, copyright, patents, trademarks, trade secrets and database rights. As such, it is the **Consortium Agreement that primarily lays down the foundation for the Intellectual Property Rights Policy, identifying Background IPR brough by partners to the project and** providing the general framework regarding, among the others, the following points:

- **a)** The **procedure to exercise partners' right to object to dissemination;**
- **b)** The **procedure of making research outputs available** to larger scientific and research communities or peer-reviewed publications only by its proprietors or through authorization;
- **c)** An **assessment of the background knowledge of project partners**, their potential contribution to the creation of foreground IP, and potential IP overlap;
- **d)** The **partners' agreement to grant the consortium with Access Rights to necessary background knowledge** for the successful execution of the project, including the conditions for sublicensing, where desired;
- **e)** The **protection of confidential information** that may be leveraged on but not published;
- **f)** The **wide dissemination of non-sensitive and not reserved for exploitation knowledge** with the EC open access policy;
- **g)** The **further use, development and exploitation of portable software components** developed in the context of the project by the consortium members;
- **h)** The **results of the termination of a partner's participation** on its obligation to grant access rights to other partners until the end of the project.

Pre-developed IPR brought into the project by partners remain their own property and are not affected by their participation to AD4GD. As such, partners are free to decide how access to such IPR may be granted, the purposes for which it may be used, as well as the duration of the access to it.

[32] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union [2016] OJ L 194/1.
[33] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) [2022] OJ L 333/80.

Where IPR owned by a third party is brought into the project, the partner introducing it must ensure that they possess the required legalization documents for their use, including licenses. Such documentation must also be shared with partners and stored in the project's private repository so that it remains accessible and available if required.

Similarly, **IPR that was generated in the context of the project is by default owned by the partner who developed the innovation or content.** Where more partners have participated in the development of the same IPR, they are required to sign a written agreement where their respective rights and obligations will be thoroughly described. Such agreement may also include the terms and conditions of IPR exploitation.

In line with the Open Science and FAIR data principles, the project will prioritize open access to its datasets, products and/or solutions, as defined by the project, aiming at outcomes that are open, standard-based and will prevent vendor lock-in to the greater extent possible and in balance with the partners' IPR generated in the context of the project.

In order to provide additional guidance to the partners, a **Data and Research Outputs Management Plan will be designed in order to maximize the value of the outcomes,** including IPR. Said plan shall encompass **guidelines and mechanisms to exploit in a sustainable manner:**

(i) Research generated datasets;
(ii) Original software developed in the project's activities;
(iii) Any new created materials, including scientific papers;
(iv) Research generated intellectual property, including licensing options.

In order to facilitate the Consortium in the management of its IPR, **an IPR Policy will be designed encompassing an adequate IPR management and strategy, also in accordance with the FAIR principles**. This IPR Policy will be signed by all partners. **A designated Dissemination Manager and the Project Coordinator will overview and monitor compliance with the IPR rules** established in order to ensure the smooth operation of the project.

Since IPR is a major part of the exploitation phase as well, the development of IPR by the partners will be closely monitored and updated accordingly in the following iterations of the Data Management Plan. If the project's IPR is infringed in any manner or form, the Consortium member identifying the infringement shall notify the Project Coordination Board immediately in order to proceed to protective measures. Accordingly, if IPR-related conflicts arise within the Consortium, partners are required to immediately notify the Project Coordination Board in order to prioritize its amicable resolution.

7. CONCLUSION AND FUTURE WORK

The current deliverable describes in the greatest extent possible the data that will be involved in the course of the project in its current stage of development, as well as the steps to ensure they are compliant with the Open Science and FAIR principles, ethical and legal provisions. It also provides further guidance on the obligations arising from the relevant legal and ethical framework in order to facilitate partners when finalizing their data management strategy and points of action.

The preliminary information reported in the present deliverable will be further expanded and complemented by the upcoming second iteration of the Data Management Plan when more detailed information on the partner's precise workplan is available. The Questionnaire provided in Annex I of the present deliverable will serve as the baseline for the partner's work and will include the elements that will be reported in this second iteration.

8. REFERENCES

- *AD4GD Project 101061001. GRANT AGREEMENT. European Research Executive Agency (REA)*
- *European Commission, 'The European Commission's priorities; 6 Commission priorities for 2019-24' (16 July 2019), available at <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024_en>.*
- *European Commission, 'A European Green Deal; Striving to be the first climate-neutral continent' (11 December 2019), available at < https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en>.*
- *GO FAIR, 'FAIR Principles' available at: < https://www.go-fair.org/wp-content/uploads/2022/01/FAIRPrinciples_overview.pdf>.*
- *UNESCO, 'UNESCO Recommendation on Open Science' (2021) <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.*
- *GEO Data Management Principles Task Force, 'GEOSS Data Management Principles' (28 April 2015).*
- *Data Sharing Working Group of the Group on Earth Observations, 'GEOSS Data Sharing Principles post 2015' (10 March 2014).*
- *WP29, 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679' (WP248 rev.01, 4 October 2017) 9-11.*
- *S Ziegler and others, Personal Data Protection for Internet of Things deployment: Lessons learned from the European Large-Scale Pilots of Internet of Things, February 2020.*
- *WP29, 'Guidelines on transparency under Regulation 2016/679' (WP260 rev.01, 11 April 2018).*
- *Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.*
- *EDPB, 'Guidelines 01/2022on data subject rights - Right of access' (version 1.0, 18 January 2022) [46].*
- *WP29, 'Guidelines on the right to data portability' (WP 242 rev.01, 5 April 2017) 9-10.*
- *EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (version 2.0, 20 October 2020) [35].*
- *European Commission, 'On Artificial Intelligence - A European approach to excellence and trust' (White Paper) COM (2020) 65 final, 2.*
- *High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (8 April 2019) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>.*
- *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L 201/37.*
- *European Commission, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' COM(2017) 10 final.*
- *Regulation (EU) 2022/868 of The European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) [2022] OJ L 152/1.*
- *European Commission, 'A European strategy for data' (Communication) COM(2020) 66 final.*
- *"Public sector bodies" are defined as "the State, regional or local authorities, bodies governed by public law or associations formed by one or more such authorities, or one or more such bodies governed by public law".*

- *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172/156.*
- *European Commission, 'Proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' COM (2022).*
- *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 077/20.*
- *European Parliament and Council of the European Union, "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information," June 20, 2019, http://data.europa.eu/eli/dir/2019/1024/oj/eng.*
- *European Commission, 'Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts' COM(2021) 206 final.*
- *Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L 303/59.*
- *Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union [2016] OJ L 194/1.*
- *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) [2022] OJ L 333/80.*

9. ANNEX I – DATA MANAGEMENT QUESTIONNAIRE

PART A - DATA SUMMARY

1. **What is the purpose of your data generation/collection/processing and how is it related to the objectives of the project?**

……
……
……

2. **What types and formats of data will you generate/collect/process within the project? Please list below.**

……
……
……

3. **Are you in charge of making decisions about what data to be collected/ processed, how and for what purpose?**
   - ☐ Yes, namely ……
   - ☐ No

4. **Do you process data under another partner's behalf/instructions?**
   - ☐ Yes, namely ……
   - ☐ No

5. **Do you currently or will you in the future share data with other partners inside the project? If yes, please list below:**
   - **to whom,**
   - **for which purpose and task,**
   - **whether the data will be anonymized/ pseunonymized or raw, and**
   - **whether you have an agreement in place with the respective party.**

   ……
   ……
   ……

6. **Will you be reusing data for further purposes beyond the project? If yes, please specify what data and for which purposes.**

……
……
……

2. Will you re-use any existing datasets? If so, please fill out the table below.

	Please provide your answers in this column:
Dataset(s) name	*What is the name of the used dataset(s)?*
Dataset(s) description	*Please provide a short description of the dataset(s).*
Personal Data	*Does the dataset include personal data? If yes, please specify the typ of personal data.*
Purpose	*What is the purpose for which you use/ process the dataset(s)?*
Data format	*What format(s) are your dataset(s)?*
Data Storage	*Where will you store the dataset(s)?*
Main Data Source	*What is the main source of the dataset(s)?*
Data Ownership	*Who owns the dataset(s)?*
Country of Origin	*Where does the dataset come from?*
Restrictions on the use	*Are there any restrictions for the use of the datasets?*
Access	*Who has access to the datasets? Please include other work package which will also access the datasets.*
Retention Period	*How long will you keep the datasets?*
Licence	*Under which licence did you obtain access to the datasets?*
WP and task	*For which work package and which task do you need to use th datasets?*
Additional Comments	*Please add here any additional comments.*

PART B – FAIR DATA

MAKING DATA FINDABLE

1. **Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism? If so, please describe.**

……
……
……

1. **What naming conventions do you follow?**

……
……
……

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

……
……
……

2. **Do you provide clear version numbers?**

……
……
……

3. **What metadata will be created, if any?**

……
……
……

MAKING DATA OPENLY ACCESSIBLE

1. **Which data produced and/or used in the project will 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.**

……
……
……

2. **How will the data be made accessible?**

……
……
……

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

……
……
……

2. **Is documentation about the software needed to access the data included? If so, please explain.**

……
……
……

3. **Is it possible to include the relevant software? Please elaborate.**

……
……
……

4. **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.**

……
……
……

5. **Have you explored appropriate arrangements with the identified repository? If so, please explain.**

……
……
……

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

……
……
……

7. **Is there a need for a data access committee? Why or why not?**

……
……
……

8. **Are there well described conditions for access (i.e. a machine readable license)? If so, please explain.**

……
……
……

1. **How will the identity of the person accessing the data be ascertained?**

……
……
……

MAKING DATA INTEROPERABLE

1. **Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, 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)?**

☐ Yes, namely ……

☐ No

2. **What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?**

……
……
……

3. **Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability? If so, please explain.**

……
……
……

4. **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?**

……
……
……

INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES)

1. **How will the data be licensed to permit the widest re-use possible?**

……
……
……

1. **Will the data be made available for re-use? If so, when? 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.**

……
……
……

2. **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.**

☐ No, because……

☐ Partially, namely …….

☐ Yes.

3. **How long is it intended that the data remains re-usable?**

……
……
……

4. **Are data quality assurance processes described?**

☐ No

☐ Yes, in ……

PART C – ALLOCATION OF RESOURCES

1. **Are there additional costs for making data FAIR in your project?**

……
……
……

2. **How will these be covered?**

……
……
……

3. **Who will be responsible for data management in your project?**

……
……
……

1. **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)?**

……
……
……

PART D – DATA SECURITY

1. **What technical and organisational measures have you put in place to ensure data security? (i.e., anonymisation techniques, pseudonymisation, tokenization, etc)**

……
……
……

2. **Is the data safely stored in certified repositories? Please describe**

……
……
……

3. **How long are you storing the data?**

……
……
……

PART E – ETHICAL ASPECTS

1. **Are there any ethical or legal issues that can have an impact on data sharing? If so, please describe.**

……
……
……

2. **Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data? If so, please describe.**

……
……
……

PART F – INTELLECTUAL PROPERTY RIGHTS

1. **Are you bringing any intellectual property (IP) to the GATEKEEPER project? If yes, please specify.**

☐ No

☐ Yes, namely ……

2. **Have you or do you foresee developing, or being involved in the development of new IP in the scope of the project? If yes, please specify.**

☐ No

☐ Yes, namely ……

3. **In case of mutual development of IP with other project's partners, what would be your requirements in regard to protection of the developed IP?**

……
……
……

4. **If the project produces IP-protection-eligible results, how will you exploit those results?**

……
……
……

5. **Will you enable free access rights (Open Access/ Open Source) after the end of the project? If so, for how long?**

……
……
……

10. ANNEX II – RELEVANT DATA PROTECTION DEFINITIONS

In order to facilitate the actors involved in the AD4GD research project, this Annex provides a table of fundamental definitions found in the European Data Protection legal framework. As previously stated, it is important to have knowledge of these basic definitions so as to assure alignment in the communication within the consortium.

MAIN DEFINITIONS	
PERSONAL DATA	According to Art. 4(1) GDPR 'personal data' means "any information relating to an identified or identifiable natural person". This definition covers any kind of statement about a living person, both objective and subjective, regardless of its correctness, and of the format or the medium on which it is contained. Anonymized data as well as data relating to legal persons are not personal data.
PROCESSING (OF PERSONAL DATA)	According to Art. 4(2) GDPR, processing of personal data "means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction."
ACTORS AND RESPONSIBILITY ROLES	
DATA SUBJECT	The data subject is an identified or identifiable natural person whose personal data are being processed. According to Art. 4 (1) GDPR, "an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Legal persons cannot assume the role of data subjects.
JOINT DATA CONTROLLERS	When two or more controllers "*jointly determine the purposes and means of processing, they shall be joint controllers*" (Art. 26 (1) GDPR). The joint determination of the purposes and means of the processing implies that more than one entity exercises a decisive influence over the identification of the reasons and modalities of the processing. This joint determination can take different forms, such as: • **A common decision**, when controllers share a common intention and decide together the key elements of the processing; • **Converging decisions**, when controllers' decisions on the means and purpose of the processing complement each other and are necessary for the processing to take place. The processing by each party shall be inextricably linked, so that it would not be possible without both parties participation.
DATA PROCESSOR	A data processor is a "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*" (Art. 4(8) GDPR). A party will act as a processor if: • It is **a separate entity** in relation to the controller, meaning that a department within a company cannot be processor to another department within the same entity, but it must be an external organisation; • It processes personal data **on the controller's behalf**, meaning that the processing is carried out for the benefit of the controller, following the

	instructions of the controller on the purposes and (essential) means of the processing. The instructions of the controller may leave room to the processor for the determination of non-essential means of the processing, such as the most suitable technical and organisational means of the processing (e.g. which type of hardware or software should be used, which detailed security measures should be adopted). A processor will qualify as a controller if it goes beyond the controller's instructions and determines its own purposes and means of processing.
DATA PROTECTION AUTHORITY	A DPA is an independent body which is in charge of: • Monitoring the processing of personal data within its jurisdiction (country, region or international organization); • Providing advice to the competent bodies with regard to legislative and administrative measures relating to the processing of personal data; • Hearing complaints lodged by citizens with regard to the protection of their data protection rights. According to Art. 51 GDPR, each Member State shall establish in its territory at least one DPA, which shall be endowed with investigative powers (such as access to data, collection of information, etc.), corrective powers (power to order the erasure of data, to impose a fine or a ban on processing, etc.), and authorisation or advisory powers (issuance of opinions, power to accredit certification bodies, etc.). As such, national data protection authorities have been established in all European countries.
PRINCIPLES	
LAWFULNESS AND FAIRNESS	The processing of personal data is lawful when the data controller has identified a legal basis for it and when applicable local laws and regulations are complied with. Art. 6 (1) GDPR provides a list of valid legal basis for data processing which can be relied upon by data controllers. According to the fairness principle, controllers shall process personal data in a way that is not unjustifiably detrimental, unexpected or misleading to data subjects.
TRANSPARENCY	The principle of transparency requires data controllers to process personal data in a transparent manner in relation to the data subject. This implies that the data controller abides by the obligation to give data subjects all required information about the processing activities they are carrying out, pursuant to Art. 13-14 GDPR.
PURPOSE LIMITATION	According to the principle of purpose limitation (Art. 5 (1)(b) GDPR), the data controller must collect data for specified, explicit, and legitimate purposes, and not further process the data in a manner that is incompatible with the purposes for which they were collected. According to Art. 5 (1)(b) GDPR, there is a presumption of compatibility between the purpose for which personal data have been originally processed and further processing for scientific research purposes, provided that the safeguards listed in Art. 89 (1) are adopted (e.g. technical and organizational measures to ensure respect for the principle of data minimisation, such as data pseudonymisation).
DATA MINIMIZATION	According to the principle of data minimization, a data controller shall process only personal data which are strictly relevant and necessary to the purpose of the processing. This implies that data controllers shall assess whether the envisaged objectives can be accomplished using less intrusive means (for instance, whether pseudonymised or anonymised data can be processed instead of personal data in plain text). This assessment shall be carried out periodically.

STORAGE LIMITATION	The principle of storage limitation requires that personal data shall be stored or kept in a form which permits the identification of data subjects for no longer that it is necessary for the specified purposes. This implies the determination of specific retention periods for the data processed for each purpose.
ACCURACY	According to Art. 5 (1)(d) GDPR, data controllers shall process personal data which are accurate, and where applicable, kept up to date. Controllers shall adopt any reasonable measure to ensure the accuracy of the data, including the erasure or the rectification of inaccurate data. These measures can relate both to the moment of collection of data and to the subsequent processing of data. It is possible that a controller keeps record of inaccurate data so to avoid incurring into the same inaccuracy at a later stage.
INTEGRITY AND CONFIDENTIALITY	Controllers and processors shall ensure that personal data are processed in a secure manner, adopting technical and organisational measures which are adequate to protect the data against unauthorised or unlawful processing and against accidental loss, destruction or damage (Art. 5 (1)(f) GDPR). Art. 32 GDPR further specifies this obligation, providing a methodology to assess which measures shall be considered appropriate. The identification of appropriate security measures should adopt a risk-based approach and should consider the state of the art and the cost of implementation. This evaluation should be carried out periodically to ensure that the measures adopted are still appropriate according to the processing activities and to technical developments.
ACCOUNTABILITY	According to the principle of accountability, data controllers shall be able to demonstrate compliance with all the aforementioned principles and with all the obligations provided under applicable data protection provisions. As a result, data controllers shall document any assessment and any technical and organizational measure undertaken with the scope of complying with those rules, so to demonstrate their appropriateness and effectiveness. Those measures shall be periodically reviewed and updated where necessary (Art. 24 (1) GDPR).
PRIVACY BY DESIGN AND BY DEFAULT	According to the principle of data protection by design, data controllers shall adopt technical and organizational measures to implement data protection principles and to protect the rights and freedoms of data subjects at the time of the determination of the means for processing (Art. 25 (1) GDPR). The determination of appropriate safeguards shall take place before the processing begins, namely, when the controller is deciding how the processing is going to be conducted and through which mechanisms. The adequacy of such measures shall be periodically assessed during the whole duration of the processing. According to the principle of data protection by default, data controllers shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. This principle therefore requires that a given technology used to process data adopts by default the most privacy-preserving method to perform it. It also requires that organizational measures adopted to support processing operations comply with the principle of data minimization (e.g. allocating data access to personnel on the basis of a need-to-know basis).
DATA SUBJECT RIGHTS	
RIGHT TO INFORMATION	Data subjects have the right to be informed about the processing of personal data concerning them. If data are collected from an individual, they shall be informed as to who is collecting their data, how to contact the controller and its data protection officer, for which purpose and on which legal grounds the data is processed, who will also receive the data, for how long it will be kept and how this period is

	determined, and whether automated decision-making is involved. This also includes receiving information on the rights available to them as well as the right to lodge a complaint with a data protection authority (Art. 13 GDPR). If data is not collected directly from the individual, they still have the right to be informed about the processing of personal data relating to them. In this case, the information shall also specify the categories of the data being processed and the source from which they originated (Art. 14 GDPR).
RIGHT TO ACCESS	The data subjects shall have the right to obtain from the controller confirmation as to whether or not personal data concerning them are being processed. If that is the case, data subjects shall have access to the personal data that are being processed in an intelligible form and be informed on: • The purposes of the processing; • The recipients to whom the personal data has been or is going to be disclosed; • The existence of the right to request rectification or deletion of the data, restriction to the processing or to object to it, as well as the right to lodge a complaint with a data protection authority; • Where possible, the envisaged period for which the personal data will be stored; • Where the personal data are not collected from the data subject, any available information as to their source; • The existence of automated decision-making and meaningful information about the logic involved. The scope of the right to access is to provide data subjects with information about the processing of their personal data, so to allow them to verify the lawfulness of the processing and to exercise other rights provided by data protection provisions. Access can be granted only to the personal data concerning the person making the request, while access to personal data relating to other persons can only be allowed subject to appropriate authorization.
RIGHT TO RECTIFICATION	The data subject has the right to obtain from the controller the rectification of inaccurate or incomplete personal data concerning them. Whether personal data is inaccurate or incomplete shall be assessed with regard to the purpose of the processing. The controller must notify any rectification of personal data to each recipient to whom the personal data has been disclosed, unless this proves impossible or involves a disproportionate effort (Art. 19 GDPR).
RIGHT TO OBJECT	In certain circumstances, data subjects may have the right to object to the processing of their personal data on the basis of their particular situation. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims (Art. 21 GDPR).
RIGHT TO RESTRICTION OF PROCESSING	According to Art. 18 GDPR, in certain circumstances, data subjects may have the right to obtain from the controller the restriction of processing of their personal data. In this case, such personal data shall, with the exception of storage, only be processed: • With the data subject's consent; • For the establishment, exercise or defence of legal claims;

	• For the protection of the rights of another natural or legal person; • For reasons of important public interest of the Union or of a Member State.
RIGHT TO DATA PORTABILITY	When the processing is performed by automated means and it is based on the data subject's consent, or is necessary for the execution of a contractual obligation, Art. 20 GDPR provides data subjects with the right to data portability. In this case data subjects have the right to receive their personal data in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller, where technically feasible. The right to data portability only applies to personal data provided by the data subject to a controller (e.g. email address) and to personal data gathered from the observation of a data subject (e.g. raw data processed by a smart object). This implies that personal data created by the controller on the basis of provided/observed data are out of the scope of this right.
RIGHT NOT TO BE SUBJECT TO AUTOMATED DECISION-MAKING	Art. 22 GDPR provides data subjects with the right not to be subject to decisions based solely on automated decision-making or profiling which creates legal or similar effects for such persons, unless the decision: • Is necessary for entering into, or performance of, a contract between the data subject and a data controller; • Is authorised by Union or Member State law to which the controller is subject; • Is based on the data subject's explicit consent. In such cases, the controller shall at least ensure that the data subject has the means to obtain human intervention, to express their point of view and to contest the decision.
KEY COMPLIANCE DOCUMENTATION	
DATA PROCESSING AGREEMENT	When a controller decides to delegate the entire or part of the processing activities to a data processors, the processing shall be governed by a contract (Data Processing Agreement) which shall have the minimum content as required by Art. 28 GDPR. The Data Processing Agreement shall stipulate at least that the data processor: • Shall act only on instructions from the data controller including with regard to transfers of personal data to a third country or an international organisation; • Ensures that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality; • Takes all measures required pursuant to Art. 32 GDPR (e.g. technical and organizational security measures); • Engages a sub-processor only with prior written authorization of the controller. A sub-processor can be appointed using a written contract which places on the sub-processor the same data protection obligations placed on the processor by the controller; • Assists the controller in ensuring compliance with the controller's obligations under the GDPR; • At the choice of the controller, when the relationship ends, deleting or returning to the controller all the personal data relating to processing and deleting existing copies unless Union or Member State law requires storage of the personal data.
DATA CONTROLLERSHIP	Art. 26 GDPR requires joint controllers to adopt an arrangement in which they shall in a transparent manner determine their respective responsibilities for compliance

AGREEMENT	with the GDPR, and reflect their respective roles and relationships vis-à-vis the data subjects. The essence of the DCA shall be made available to data subjects.
DATA PROTECTION IMPACT ASSESSMENT	According to Art. 35 GDPR, when the processing of personal data is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. In doing so, the controller shall seek the advice of the data protection officer, where designated. Each national DPA has adopted a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment.