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: D8.3

DATA MANAGEMENT PLAN

Due date of deliverable:	29/02/2024
Actual submission date:	29/02/2024
Version:	3
Main Authors:	Ekaterina Kasyanova-Kühl, 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	D8.3
Deliverable version	3
Contractual date of delivery	29/02/2024
Actual date of delivery	29/02/2024
Document status	Final
Document version	1
Dissemination	Public
Work package	WP8: Project Management and Coordination
Partner responsible	European Centre for Certification and Privacy (ECCP)
Author(s)	Ekaterina Kasyanova-Kühl, Vasiliki Tsiompanidou, Claudia Martorelli, Adrian Quesada Rodriguez, Renata Radocz, Ana Maria Pacheco, Sébastien Ziegler, Stefan Schiffner
Editor(s)	Ekaterina Kasyanova-Kühl, Joan Maso Pau
Reviewer(s)	Joan Maso Pau
EC Project Officer	Lara Congiu

Abstract	The Data Management Plan provides guidance and documentation on data management and personal data protection principles that need to be followed in the course of the project. It explains how partners are managing their data 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 regular-li reviewed and updated throughout the project's lifecycle reflecting how the partners are implementing the initial strategy into managing practices at different points of the project. In this intermediate report, we collected practices for most of the datasets in the project and ensured that the information contained in this document is consistent with the status of the project data catalogue (GeoNetwork). There are still some gaps in some aspects of the management that will be filled during the second part of the project lifespan.
Keywords	Data management, data protection, privacy, data security, FAIR data, intellectual property rights.
Disclaimer	

DOCUMENT VERSION HISTORY

Version history			
Version	**Date**	**Modification reason**	**Modified by**
1	26/01/2024	Initial version of the document based on the preceding iteration of the Data Management Plan (D.8.2) submitted in February 2023	Ekaterina Kasyanova-Kühl (ECCP)
2	19/02/2024	Internal review	Joan Maso Pau (CREAF)
3	20/02/2024	Addressed comments following the internal review	Ekaterina Kasyanova-Kühl (ECCP) Adrian Quesada Rodriguez (MI)
4	29/02/2024	Final edits	Joan Maso Pau (CREAF)

ABBREVIATIONS

Abbreviation	Definition
AD4GD	AllData4GreenDeal, the current project
AI	Artificial Intelligence
API	Application Programming Interface
CitSci	Citizen Science
CFREU	Charter of Fundamental Rights of the EU
DCA	Data Controllership Agreement
DGA	Data Governance Act
DPA	Data Protection Authority
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EC	European Commission
EDPB	European Data Protection Board
EOSC	European Open Science Cloud
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Re-usable
GDPR	General Data Protection Regulation (Regulation (EU) 2016/679)
GEOSS	Global Earth Observation System of Systems
IoT	Internet of Things
IPR	Intellectual Property Rights

NIS	Network and Information Security Directive
TOM	Technical and Organisational Measures
UN	United Nations
WP	Work Package
WP29	Article 29 Working Party

Table of Contents

List of Figures

List of Tables

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. Only 2 partners declared to dela with personal information but none of them provided any details about their datasets. 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. Nine datasets provided by three partners are described in this version of the DMP.

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

1 INTRODUCTION

A Data Management Plan (DMP) is a key element of good data management. The DMP describes the data management life cycle for the data to be collected, processed and/or generated by this project. The DMP should describe the efforts done in the project to make data more findable, accessible, interoperable and re-usable (FAIR). In addition, a DMP includes information on: the handling of research data during & after the end of the project, what data will be collected, processed and/or generated, which methodology & standards will be applied, whether data will be shared/made open access and how data will be curated & preserved (including after the end of the project). A DMP is required for all projects participating in Horizon Europe.

1.1 OBJECTIVES OF THE PROJECT

In the past decade, environmental and sustainability concerns have become a bigger 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 is 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 achieving climate-neutrality in the EU by 2050[2].

A major part of the European 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, the main objective of the AD4GD 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 intends to 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.

[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 21 January 2024].

[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 21 January 2024].

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.
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.

1.2 PURPOSE OF THE DOCUMENT

The present document intends to illustrate the various categories of data that are collected/processed/generated or are intended to be collected/processed/generated in the course of the project, whether falling under the category of personal data or not. In this context, the document will focus on providing 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 is or 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.

1.3 METHODOLOGY

The present Data Management Plan explores the implications of the handling of personal and non-personal data, 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 existing as well as anticipated legislation on the European level. In addition, the requirements for data to be FAIR are analysed and the project's and partners' initial approach towards FAIR data is explained. Finally, this DMP describes the Consortium's existing and upcoming strategy regarding the management of Intellectual Property Rights.

2 RELEVANT STANDARDS AND PRINCIPLES FOR THE DATA PROCESSING IN AD4GD

2.1 OVERVIEW

In line with the European Strategy for Data[3], 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 not only collects and/or processes data, but also to generates a number of datasets, as will be further analysed and showcased 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 compliant with data protection and privacy obligations, 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[4], 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.**

The characteristics of datasets that are already generated/collected/processed in the course of the project have been determined through the DMP questionnaire (Annex I) that was distributed among the partners. The questionnaire consists of six parts:

1. **Part A – Dataset Description.** In this section, the partners were called to describe a dataset that is used, generated, collected and/or processed within the context of AD4GD.
2. **Part B – Data Summary**. This refers to the categories of data intended to be managed (created or processed).
3. **Part C – FAIR Data**. This section required partners to explain the measures implemented to make the dataset adhere to FAIR principles.

[3] European Commission, 'Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data' (19 February 2020), COM(2020) 66 final.
[4] UNESCO, 'UNESCO Recommendation on Open Science' (2021) <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.

4. **Part D - Allocation of Resources**. In this section, the partners were asked to indicate any cost implications that making a dataset FAIR would entail.
5. **Part E** - **Ethical and Legal Aspects**. This section refers to any ethical or legal issues identified by partners organisations that have impacted or might impact the dataset sharing.
6. **Part F – Intellectual Property Rights**. In this section, the partners were called to explain the Intellectual Property Rights (IPR) brought to the project or generated through it.

The results of the questionnaire are reported in detail in the sections below. Using the said questionnaire, partners were able to better reflect on, build and report on their data generation and processing activities, their data management strategy, as well as their IPR-related exploitation ambitions. As the project progress, the datasets are expected to evolve, thus certain characteristics of the datasets might be extended and further elaborated during the later stage of the project. The updates, where applicable, will be reported in the next iteration of the Data Management Plan.

2.1.1 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 generated 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 organisations such as WMO, GEO, and the UN system.

Provided that (as described below) both personal and non-personal is involved, AD4GD adopts a strategy, based on both ensuring privacy by design and by default when personal data is concerned and openness, availability, interoperability, accessibility, and reusability with regard to non-personal data.

2.1.2 CATEGORIES OF DATA

Within the AD4GD project, various datasets are processed and generated, encompassing different categories of data. These datasets are intended to serve specific purposes and contribute to the project's objectives. As already explained, with the progress of the project, the datasets are expected to evolve. Nevertheless, following the results of the conducted DMP questionnaire, it is already possible to identify the relevant data for AD4GD:

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 non-personal 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.

Table 1 illustrates the types of data managed (created or processed) by each AD4GD partner.

Table 1: Categories of data intended to be managed (created or processed) by AD4GD partners

Partner	Dataset name (if applicable)	Type of data
Fraunhofer Institute for Applied Information Technology	N/A	Personal data

CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	Non-personal data
ASTON	Interaction Flow (IF) index	Non-personal data
ECMWF	Sensor Community Air Quality data	Non-personal data
KWB	Crowdwater Data	Non-personal data
	DWD Weather Data	Non-personal data
	Berliner Wasserportal: Water level	Non-personal data
	Berliner Wasserportal: Water temperature	Non-personal data
	KWB: Oxygen concentration	Non-personal data
	Sentinel-2	Non-personal data
	KWB: Trophic state estimation	Non-personal data
IoT Lab	N/A	Personal data Non-personal data

Where personal data is processed or is intended to be processed, the further analysis on a per-partner basis should be provided in a specific section. Currently the two partners that deal with personal information did not provided information about any dataset in this data management plan. In practice, this mean that this draft of the DMP only describes datasets of non-personal information.

In order to further determine the datasets and their specifications and characteristics, the following table has been provided in Part A of the questionnaire to be filled out by all partners.

Table 2: Part A of the DMP questionnaire on dataset description

Question	Response
Name of the used dataset(s)	*To be provided by the partner*
Reference to GeoNetwork record	*To be provided by the partner*
Short description of the dataset(s)	*To be provided by the partner*
Purpose for which you use/ process the dataset(s)	*To be provided by the partner*
Format(s) of dataset(s)	*To be provided by the partner*
Where will you store the dataset(s)?	*To be provided by the partner*
What is the main source of the dataset(s)?	*To be provided by the partner*
Who owns the dataset(s)?	*To be provided by the partner*

Origin of the dataset	*To be provided by the partner*
Are there any restrictions for the use of the datasets?	*To be provided by the partner*
Who has access to the datasets?	*To be provided by the partner*
Under which licence did you obtain access to the datasets?	*To be provided by the partner*
Additional comments	*To be provided by the partner*

The current analysis is carried-out on a per-partner basis. Three partners from AD4GD's consortium, i.e., CREAF, ECMWF, and KWB have generated and reported the datasets as follows.

Table 3: CREAF - MUCSC dataset description – Pilot BioConn

Partner: CREAF **Pilot Name: *BioConn***	
Question	**Response**
Name of the used dataset(s)	*Land Use Land Cover Maps from Catalonia (MUCSC)*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*https://catalogue.grumets.cat/geonetwork/ad4gd/cat/catalog.search#/metadata/a110782b-b213-4922-988a-8778a66f8d9d*
Short description of the dataset(s)	*Land use and land cover mapping of Catalonia obtained through automatic classification of Landsat and Sentinel-2 images with a spatial resolution of 30 and 10 m, and with auxiliary information provided by Generalitat de Catalunya*
Purpose for which you use/ process the dataset(s)	*Input for the Terrestrial Connectivity Index Maps calculation*
Format(s) of dataset(s)	*IMG (MiraMon native format) and COG*
Where will you store the dataset(s)?	*Locally and in the BioConn Open Data Cube*
What is the main source of the dataset(s)?	*Landsat, Sentinel-2 images and auxiliary GIS information*
Who owns the dataset(s)?	*Generalitat de Catalunya*

Origin of the dataset	*Created by CREAF under a contract with the Generalitat*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*CC-BY*
Additional comments	*This dataset it's a time series. There is one dataset for the following years: 1987, 1992, 1997, 2002, 2007, 2012, 2017, 2022*

Table 4: ASTON - Interaction Flow (IF) index – Pilot BioConn

Partner: ASTON **Pilot Name: *BioConn***	
Question	**Response**
Name of the used dataset(s)	*Interaction Flow (IF) index*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*https://catalogue.grumets.cat/geonetwork/ad4gd/cat/catalog.search#/metadata/95b81008-d73d-4082-9572-0c471b6f5ced*
Short description of the dataset(s)	*Connectivity of habitats of rare and endangered species in Catalonia obtained through Land-Use/Land-Cover maps (LULC) processing via Graphab open-access software and based on MUCSC data for Catalonia and user-defined values of habitat suitability (maximum distance for indefinite species to migrate - 600 meters, habitat codes are 17,18,19)).*
Purpose for which you use/ process the dataset(s)	*Have estimations of Terrestrial Connectivity Index Maps calculation*
Format(s) of dataset(s)	*N/A*
Where will you store the dataset(s)?	*Locally*
What is the main source of the dataset(s)?	*LULC maps and Graphab*

Who owns the dataset(s)?	*ASTON*
Origin of the dataset	*Created by ASTON in the project*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*TBD*
Under which licence did you obtain access to the datasets?	*TBD*
Additional comments	*This dataset it's a time series. There is one dataset for the following years: 1987, 1992, 1997, 2002, 2007, 2012, 2017, 2022*

Table 5: ECMWF - Sensor.Community Air Quality dataset description – Pilot Air Quality

Partner: ECMWF **Pilot Name: Air Quality**	
Question	**Response**
Name of the used dataset(s)	*Sensor.Community Air Quality data*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*https://catalogue.grumets.cat/geonetwork/ad4gd/eng/catalog.search#/metadata/1790de06-d90e-4824-a3b2-f4d6105db1ea*
Short description of the dataset(s)	*Sensor.Community is a collaborative network of volunteers who use DIY sensors to monitor environmental data, like air and noise pollution. These volunteers install sensors in various locations and share the collected data publicly, contributing to a better understanding of environmental issues at a local and global level.* *So far, the dataset includes time series observations of particulate matter (i.e. PM2.5 and PM10) from various IoT sensors.*
Purpose for which you use/ process the dataset(s)	*Investigations of the usefulness of IoT measurements for air quality monitoring*

Format(s) of dataset(s)	*CSV*
Where will you store the dataset(s)?	*The dataset is currently publicly available at archive.sensor.community and we will also store a copy in an internal ECMWF database.*
What is the main source of the dataset(s)?	*https://archive.sensor.community/*
Who owns the dataset(s)?	*The dataset is freely available via the platform's website.*
Origin of the dataset	*https://archive.sensor.community/*
Are there any restrictions for the use of the datasets?	*There are no access restrictions.*
Who has access to the datasets?	*Anyone can access the data.*
Under which licence did you obtain access to the datasets?	*Database Contents License (DbCL) v1.0*

KWB has provided information on seven separate datasets.

Table 6: KWB - Crowdwater Data dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*Crowdwater Data*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*TBA*
Short description of the dataset(s)	*Data for either a virtual or a physical staff gauge or lake properties, gathered by the crowdwater app for Berlin lakes*
Purpose for which you use/ process the dataset(s)	*Data will be used to evaluate water quality and availability of lakes*
Format(s) of dataset(s)	*CSV*

Where will you store the dataset(s)?	*Locally, GeoNetwork (Zenodo)*
What is the main source of the dataset(s)?	*Produced by Citizens via Crowdwater App*
Who owns the dataset(s)?	*Hydrology and Climate group* *Department of Geography* *University of Zurich* *Winterthurerstrasse 190* *8057 Zürich*
Origin of the dataset	*https://crowdwater.ch/en*
Are there any restrictions for the use of the datasets?	*We are happy if you write us a message to info@crowdwater.ch if you use the data, so we can see what can be achieved with it.*
Who has access to the datasets?	*Public*

Table 7: KWB - DWD Weather Data dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*DWD Weather Data*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*TBA*
Short description of the dataset(s)	*Precipitation, air temperature, for the weather station Berlin Tempelhofe, run by the German Weather Forecast (DWD)*
Purpose for which you use/ process the dataset(s)	*Data will be used as additional model input variables for water quality and water availability*
Format(s) of dataset(s)	*JSON*

Where will you store the dataset(s)?	*Data will be regularly downloaded and storted locally, the final data set will also be stored on GeoNetwork*
What is the main source of the dataset(s)?	*German Weather Forecast*
Who owns the dataset(s)?	*Deutscher Wetterdienst* *Frankfurter Straße 135* *63067 Offenbach*
Origin of the dataset	*https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/precipitation/historical/* *Downloaded via Brightsky API:* *https://brightsky.dev/docs/#/operations/getWeather*
Are there any restrictions for the use of the datasets?	*"all spatial data and spatial data services available for free access may be used without any restrictions provided that the source is acknowledged"*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*https://www.dwd.de/EN/service/copyright/copyright_node.html*
Additional comments	*Online Data, updated daily*

Table 8: KWB - Berliner Wasserportal: Water level dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*Berliner Wasserportal: Water level*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*https://catalogue.grumets.cat/geonetwork/ad4gd/cat/catalog.search#/metadata/fe29687e-59ab-4cde-a8ff-*

	8fa62c76bcc7_TMP0002T (partial example)
Short description of the dataset(s)	*Water level in Berlin medium sized lakes*
Purpose for which you use/ process the dataset(s)	*Input to set up a water balance*
Format(s) of dataset(s)	*CSV*
Where will you store the dataset(s)?	*Locally*
What is the main source of the dataset(s)?	*https://wasserportal.berlin.de/start.php*
Who owns the dataset(s)?	*Senatsverwaltung für Mobilität, Verkehr, Klimaschutz und Umwelt* *Am Köllnischen Park 3* *10179 Berlin*
Origin of the dataset	*https://wasserportal.berlin.de/start.php*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*https://creativecommons.org/licenses/by/4.0/deed.en*
Additional comments	*Continuous update of dataset by online sensors*

Table 9: KWB – Berliner Wasserportal: Water temperature dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	Berliner Wasserportal: Water temperature
Category of data processed	*Non-personal*

Reference to GeoNetwork record	*TBA*
Short description of the dataset(s)	*Water temperature in Berlin medium sized lakes*
Purpose for which you use/ process the dataset(s)	*Input to set up a water balance*
Format(s) of dataset(s)	*CSV*
Where will you store the dataset(s)?	*Locally*
What is the main source of the dataset(s)?	*https://wasserportal.berlin.de/start.php*
Who owns the dataset(s)?	*Senatsverwaltung für Mobilität, Verkehr, Klimaschutz und Umwelt* *Am Köllnischen Park 3* *10179 Berlin*
Origin of the dataset	*https://wasserportal.berlin.de/start.php*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*https://creativecommons.org/licenses/by/4.0/deed.en*
Additional comments	*Continuous update of dataset by online sensors*

Table 10: KWB – Oxygen concentration dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*KWB: Oxygen concentration*
Category of data processed	*Non-personal*

Reference to GeoNetwork record	*TBA*
Short description of the dataset(s)	*Oxygen concentration in small Berlin lakes measured by sensors of KWB*
Purpose for which you use/ process the dataset(s)	*Input data to derive water quality in terms of nutrients and trophic state*
Format(s) of dataset(s)	*TBA*
Where will you store the dataset(s)?	*Locally*
What is the main source of the dataset(s)?	*Produced by KWB*
Who owns the dataset(s)?	*KWB Kompetenzzentrum Wasser Berlin gGmbH* *Cicerostr. 24* *10709 Berlin*
Origin of the dataset	*No other origin*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*https://creativecommons.org/licenses/by/4.0/legalcode*

Table 11: KWB – Sentinel-2 dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*Sentinel-2*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*TBA*

Short description of the dataset(s)	*Reflection of different wavelength according to ESA-Sentinel 2 mission over Berlin medium sized lakes*
Purpose for which you use/ process the dataset(s)	*Data will be used as model input to estimate the trophic state of medium sized lakes*
Format(s) of dataset(s)	*CSV*
Where will you store the dataset(s)?	*Data will be regularly downloaded and stored locally, the final data set will also be stored on GeoNetwork*
What is the main source of the dataset(s)?	*European Space Agency*
Who owns the dataset(s)?	*https://www.esa.int/Services/Contacts*
Origin of the dataset	*ESA Copernicus Data Space Ecosystem*
Are there any restrictions for the use of the datasets?	*"All spatial data and spatial data services available for free access may be used without any restrictions provided that the source is acknowledged"*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*Creative Commons CC BY-SA 3.0 https://open.esa.int/image-usage-creative-commons/*
Additional comments	*Online Data, constantly updated*

Table 12: KWB - Trophic state estimation dataset description – Pilot Berlin

Partner: KWB **Pilot Name: Berlin**	
Question	**Response**
Name of the used dataset(s)	*KWB: Trophic state estimation*
Category of data processed	*Non-personal*
Reference to GeoNetwork record	*TBA*

Short description of the dataset(s)	*The trophic state is a model output expressed as trophic index between 1 and 5*
Purpose for which you use/ process the dataset(s)	*Sentinal-2 data will be used to derive the trophic state of medium sized lakes in Berlin*
Format(s) of dataset(s)	*TBA*
Where will you store the dataset(s)?	*Locally, GeoNetworks*
What is the main source of the dataset(s)?	*Produced by KWB*
Who owns the dataset(s)?	*KWB Kompetenzzentrum Wasser Berlin gGmbH* *Cicerostr. 24* *10709 Berlin*
Origin of the dataset	*No other origin*
Are there any restrictions for the use of the datasets?	*You must give appropriate credit, provide a link to the license, and indicate if changes were made.*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	https://creativecommons.org/licenses/by/4.0/legalcode

Table 13: PSNC – Essential Variables lists

Partner: PSCS **Pilot Name: N/A**	
Question	**Response**
Name of the used dataset(s)	*Essential Variables lists*
Category of data processed	*Non-personal*
Reference to GitHub record	*https://github.com/AD4GD/essentialVariables*
Short description of the dataset(s)	*Essential Variables List*

Purpose for which you use/ process the dataset(s)	*Semantic tagging*
Format(s) of dataset(s)	*CSV, turttle*
Where will you store the dataset(s)?	*GitHub*
What is the main source of the dataset(s)?	*Web pages with EV definitions*
Who owns the dataset(s)?	*GCOS, GEOBON, GEOGLAM*
Origin of the dataset	*Web pages with EV definitions*
Are there any restrictions for the use of the datasets?	*None*
Who has access to the datasets?	*Public*
Under which licence did you obtain access to the datasets?	*CC0*

Some of the datasets as described in Tables 3-13 are intended to be used to develop algorithmic models that aim to enhance existing Earth observation models. Some others are already results of these models.

Additionally, the project will incorporate data from existing networks of in-situ observatories such as the ENVRI plus network (Environmental Research Infrastructures Providing Shared Solutions for Science and Society), the European Citizen Science association, and the INSPIRE (Innovation in Science Pursuit for Inspired Research) community. The integration of in-situ observations, socioeconomic data, and CitSci Data will further enable the design and cross-validation of alternative models for generating environmental metrics. This holistic approach aims to foster a comprehensive understanding of the environmental factors under consideration. Satellite and Earth Observations will be utilised to determine datasets suitable for testing the proposed semantic model. Through a multi-actor approach involving relevant stakeholders and users, these datasets will facilitate rigorous examination and validation within the context of the project's case studies.

3 FAIR DATA

AD4GD ambitions to not only support but also play a catalytical role in the design of the European Green Data Space. In that sense, AD4GD data and services simulate the same problems than the GDDS will face at a different scale. FAIR data is essential for the project's goals and aspirations as well as for the GDDS. FAIR principles offer a holistic approach to enhance the value and impact of research data. By making data findable, accessible, interoperable, and reusable, researchers and institutions can foster collaboration, innovation, and scientific progress while promoting transparency, reproducibility, and knowledge-sharing.

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[5]. Further analysis follows is provided based on the partners' responses to the corresponding section of the DMP questionnaire (Section C).

3.1 MAKING DATA FINDABLE

In order to make data and metadata findable, it is essential that it meets the following conditions[6]:

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

As such, AD4GD has already recognised 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 standardised 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.

The AD4GD project has deployed a GeoNetwork metadata catalogue instance to register the metadata of all datasets in the project pilots. Geonetwork is an open-source geospatial catalogue application used for managing, discovering, and sharing geospatial metadata. It facilitates the creation of metadata catalogues that describe geospatial resources, such as datasets, services, and other geographical information. The primary goal of GeoNetwork is to improve discoverability of geospatial information by providing a standardised and interoperable platform. The URL of the Geonetwork is: https://catalogue.grumets.cat/geonetwork/ad4gd. Some record had been made available and some others remain private until the content is mature.

The table below summarises commitment of the partners to ensure data findability.

Table 14: Summary of partners' commitment to ensure data findability

Partner	Dataset	Are the data produced and/or used in the project discoverable with metadata, identifiable and

[5] GO FAIR, 'FAIR Principles' available at: < https://www.go-fair.org/wp-content/uploads/2022/01/FAIRPrinciples_overview.pdf> [accessed 30 January 2024].
[6] Ibid no 4.

		locatable by means of a standard identification mechanism?
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes* *The dataset used in AD4GD will be made available in the Catalan Data Cube and complemented by metadata according to the ISO 19115:2003(E)*
ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Yes* *As the pilot is currently in very early stages it would be too early to try to adhere fully to FAIR principles. It is not currently clear if the project will produce datasets that will be publicly shared. However, if such datasets are produced we will endeavour to make sure they have proper metadata and otherwise fulfil the FAIR criteria.*
KWB	Crowdwater Data	*Yes* *The dataset used in AD4GD will be uploaded in GeoNetwork and complemented by metadata according to the ISO 19115:2003(E)*
KWB	DWD Weather Data	*Yes* *The dataset used in AD4GD will be uploaded in GeoNetwork and complemented by metadata according to the ISO 19115:2003(E)*
KWB	Berliner Wasserportal: Water level	*Yes* *The data is already publicly available and accessible via API: Link* *The dataset used in AD4GD will be uploaded in GeoNetwork and complemented by metadata according to the ISO 19115:2003(E)*
KWB	Berliner Wasserportal: Water temperature	*Yes* *The data is already publicly available and accessible via API: Link* *GeoNetwork and complemented by metadata according to the ISO 19115:2003(E)*
KWB	KWB: Oxygen concentration	*Yes* *The dataset used in AD4GD will be uploaded in GeoNetwork and complemented by metadata according to the ISO 19115:2003(E).*
KWB	Sentinel-2	*Yes* *The dataset used in AD4GD will be uploaded in GeoNetwork and complemented by metadata according to the ISO 19115:2003(E)*
KWB	KWB: Trophic state estimation	*Yes*

		Documentation according to GeoNetwork metadata standards
Partner	**Dataset**	**Naming conventions that are followed**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*N/A*
ASTON	Interaction Flow (IF) index	*N/A*
ECMWF	Sensor Community Air Quality data	*N/A*
KWB	Crowdwater Data	*Whenever possible vocabularies for the water sector will be used*
KWB	DWD Weather Data	*Whenever possible vocabularies for the water sector will be used*
KWB	Berliner Wasserportal: Water level	*Whenever possible vocabularies for the water sector will be used*
KWB	Berliner Wasserportal: Water temperature	*Whenever possible vocabularies for the water sector will be used*
KWB	KWB: Oxygen concentration	*Whenever possible vocabularies for the water sector will be used*
KWB	Sentinel-2	*Whenever possible vocabularies for the water sector will be used*
KWB	KWB: Trophic state estimation	*Whenever possible vocabularies for the water sector will be used*
Partner	**Dataset**	**Will search keywords be provided that optimise possibilities for re-use?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes*
ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Yes*
KWB	Crowdwater Data	*Yes*
KWB	DWD Weather Data	*Yes*
KWB	Berliner Wasserportal: Water level	*Yes*
KWB	Berliner Wasserportal: Water temperature	*Yes*
KWB	KWB: Oxygen concentration	*Yes*
KWB	Sentinel-2	*Yes*
KWB	KWB: Trophic state estimation	*Yes*
Partner	**Dataset**	**Are clear version numbers provided?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*No*

ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Yes*
KWB	Crowdwater Data	*Yes*
KWB	DWD Weather Data	*Yes*
KWB	Berliner Wasserportal: Water level	*Yes*
KWB	Berliner Wasserportal: Water temperature	*Yes*
KWB	KWB: Oxygen concentration	*Yes*
KWB	Sentinel-2	*Yes*
KWB	KWB: Trophic state estimation	*Yes*
Partner	**Dataset**	**What metadata will be created, if any?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Metadata necessary for GeoNetwork template*
ASTON	Interaction Flow (IF) index	*Metadata necessary for GeoNetwork template*
ECMWF	Sensor Community Air Quality data	*N/A*
KWB	Crowdwater Data	*Metadata necessary for GeoNetwork template*
KWB	DWD Weather Data	*Metadata necessary for GeoNetwork template*
KWB	Berliner Wasserportal: Water level	*Metadata necessary for GeoNetwork template*
KWB	Berliner Wasserportal: Water temperature	*Metadata necessary for GeoNetwork template*
KWB	KWB: Oxygen concentration	*Metadata necessary for GeoNetwork template*
KWB	Sentinel-2	*Metadata necessary for GeoNetwork template*
KWB	KWB: Trophic state estimation	*Metadata necessary for GeoNetwork template or other reporting formats*

In addition to the above, where applicable, partners ensure that clear version numbers and key words are provided to optimise possibilities of the reuse and enhance findability.

3.2 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 authorisation has been concluded. In order for data to be considered accessible, it needs to meet the following criteria[7]:

i. It is **retrievable by its identifier using a standardized communication protocol;**
ii. The protocol used is **open, free, and universally implementable;**

[7] Ibid no 4.

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**.

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. AD4GD has adopted a dual approach to data accessibility:

- For small datasets that can be downloaded as files, we use the GeoNetwork "attach" mechanism that allow for uploading data to GeoNetwork and link it as attachments to the metadata.
- For bigger datasets, they are made accessible as web services for visualization and downloading. The Catalan Data Cube is one possible web service available. The service access point will be linked from the GeoNetwork metadata record.

Some data will be uploaded on a MinIO instance for internal processing purposes and made available to the rest of the project participants.

The table below summarises commitment of the partners to ensure data accessibility.

Table 15: Summary of partners' commitment to ensure data accessibility

Partner	Dataset	Which data produced and/or used in the project will be made openly available as the default
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*They are openly available*
ASTON	Interaction Flow (IF) index	*tbd*
ECMWF	Sensor Community Air Quality data	*tbd*
KWB	Crowdwater Data	*This data set will be completely openly available*
KWB	DWD Weather Data	*This data set will be completely openly available*
KWB	Berliner Wasserportal: Water level	*This data set will be completely openly available*
KWB	Berliner Wasserportal: Water temperature	*This data set will be completely openly available*
KWB	KWB: Oxygen concentration	*If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.* *This data set will be completely openly available*
KWB	Sentinel-2	*If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.* *This data set is a section of the already openly available Sentinel-2 data and will also be openly*

		available
KWB	KWB: Trophic state estimation	*If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.* *This data set will be completely openly available*
Partner	**Dataset**	**How will the data be made accessible?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Via API an open data cube or rasdaman instance.*
ASTON	Interaction Flow (IF) index	*tbd*
ECMWF	Sensor Community Air Quality data	*If datasets are produced they will be made available through ECMWFs existing data ecosystem.*
KWB	Crowdwater Data	*Data will be uploaded into the AD4GD GeoNetwork in csv format, and to IoTlab frost server*
KWB	DWD Weather Data	*Data will be uploaded into the AD4GD Database (MinIO) and GeoNetworks*
KWB	Berliner Wasserportal: Water level	*Via API or website: https://wasserportal.berlin.de/start.php and in the AD4GD GeoNetwork*
KWB	Berliner Wasserportal: Water temperature	*Via API or website: https://wasserportal.berlin.de/start.php and in the AD4GD GeoNetwork*
KWB	KWB: Oxygen concentration	*Data will be uploaded into the AD4GD GeoNetwork in csv format, and to IoTlab frost server*
KWB	Sentinel-2	*Data will be uploaded into the AD4GD Database (MinIO) and GeoNetworks*
KWB	KWB: Trophic state estimation	*The dataset used in AD4GD will be uploaded in the AD4GD database (minIO) and GeoNetworks*
Partner	**Dataset**	**What methods or software tools are needed to access the data?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Browser*
ASTON	Interaction Flow (IF) index	*Browser*
ECMWF	Sensor Community Air Quality data	*There are various API and libraries available to access ECMWF data, namely the MARS client, the Copernicus Climate Data Store API, climetlab and earthkit.*
KWB	Crowdwater Data	*Browser*
KWB	DWD Weather Data	*N/A*
KWB	Berliner Wasserportal: Water level	*Browser*
KWB	Berliner Wasserportal: Water temperature	*Browser*

KWB	KWB: Oxygen concentration	*Browser*
KWB	Sentinel-2	*N/A*
KWB	KWB: Trophic state estimation	*Browser*
Partner	**Dataset**	**Is documentation about the software needed to access the data included**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*N/A*
ASTON	Interaction Flow (IF) index	*No*
ECMWF	Sensor Community Air Quality data	*Depending on which portals are used the make the data available, this documentation is already online.*
KWB	Crowdwater Data	*No*
KWB	DWD Weather Data	*No*
KWB	Berliner Wasserportal: Water level	*No*
KWB	Berliner Wasserportal: Water temperature	*No*
KWB	KWB: Oxygen concentration	*No*
KWB	Sentinel-2	*No*
KWB	KWB: Trophic state estimation	*No*
Partner	**Dataset**	**Is it possible to include the relevant software?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*N/A*
ASTON	Interaction Flow (IF) index	*N/A*
ECMWF	Sensor Community Air Quality data	*Yes*
KWB	Crowdwater Data	*N/A*
KWB	DWD Weather Data	*N/A*
KWB	Berliner Wasserportal: Water level	*N/A*
KWB	Berliner Wasserportal: Water temperature	*N/A*
KWB	KWB: Oxygen concentration	*N/A*
KWB	Sentinel-2	*N/A*
KWB	KWB: Trophic state estimation	*N/A*
Partner	**Dataset**	**Where will the data and associated metadata, documentation and code be deposited?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*GeoNetwork: Link*
ASTON	Interaction Flow (IF) index	*GeoNetwork: Link*
ECMWF	Sensor Community Air Quality data	*The data and associated metadata will be accessible*

		through the ECMWF ecosystems and. Once the data is published, the associated documentation will be available online (tbd at a later stage of the project). *The metadata and the associated documentation will also be made available on GeoNetwork: Link.*
KWB	Crowdwater Data	*GeoNetwork: Link*
KWB	DWD Weather Data	*GeoNetwork: Link*
KWB	Berliner Wasserportal: Water level	*GeoNetwork: Link*
KWB	Berliner Wasserportal: Water temperature	*GeoNetwork: Link*
KWB	KWB: Oxygen concentration	*Preference should be given to certified repositories which support open access where possible.* *GeoNetwork: Link*
KWB	Sentinel-2	*Preference should be given to certified repositories which support open access where possible.* *GeoNetwork: Link*
KWB	KWB: Trophic state estimation	*Preference should be given to certified repositories which support open access where possible.* *GeoNetwork: Link*
Partner	**Dataset**	**If there are restrictions on use, how will access be provided?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*N/A*
ASTON	Interaction Flow (IF) index	*No restrictions*
ECMWF	Sensor Community Air Quality data	*No restrictions for Sensor Community data*
KWB	Crowdwater Data	*No restrictions*
KWB	DWD Weather Data	*No restrictions*
KWB	Berliner Wasserportal: Water level	*No restrictions*
KWB	Berliner Wasserportal: Water temperature	*No restrictions*
KWB	KWB: Oxygen concentration	*No restrictions*
KWB	Sentinel-2	*No restrictions*
KWB	KWB: Trophic state estimation	*No restrictions*
Partner	**Dataset**	**How will the identity of the person accessing the data be ascertained?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*The identity of the person accessing data is not needed*
ASTON	Interaction Flow (IF) index	*The identity of the person accessing data is not needed*
ECMWF	Sensor Community Air Quality data	*ECMWFs data ecosystem has an authentication mechanism that ensures this information is known.*

KWB	Crowdwater Data	*The identity of the person accessing data is not needed*
KWB	DWD Weather Data	*The identity of the person accessing data is not needed*
KWB	Berliner Wasserportal: Water level	*The identity of the person accessing data is not needed*
KWB	Berliner Wasserportal: Water temperature	*The identity of the person accessing data is not needed*
KWB	KWB: Oxygen concentration	*The identity of the person accessing data is not needed*
KWB	Sentinel-2	*The identity of the person accessing data is not needed*
KWB	KWB: Trophic state estimation	*The identity of the person accessing data is not needed*

3.3 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[8]:

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.

The project has a full WP devoted to semantic interoperability on the GDDS (WP1). The vocabularies developed during the project will be applied to the different data assets in the project in different manner. Tools, such as the OGC rainbow (formerly known as OGC definition server), the AD4GD information model, the meaning catalogue, the meaning service and TAPIS will help in this direction. AD4GD 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

[8] Ibid no 4.

ontologies and vocabularies already available for the targeted domains (e.g. EIONET, QUDT, etc), exploiting as much as possible existing INSPIRE, OGC and other standards, while extending them.

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.

All these approaches will be tested in the datasets of the project. The table below summarises commitment of the partners to ensure data interoperability.

Table 16: Summary of partners' commitment to ensure data interoperability

Partner	**Dataset**	**Are the data produced in the project interoperable?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes, namely, metadata standards and naming conventions*
ASTON	Interaction Flow (IF) index	*Yes, namely, metadata standards and naming conventions*
ECMWF	Sensor Community Air Quality data	*Yes, if datasets are produced they will be made available in standard scientific formats.*
KWB	Crowdwater Data	*Yes, namely, metadata standards and naming conventions*
KWB	DWD Weather Data	*Yes, namely, STA implementation and use of metadata standards for documentation*
KWB	Berliner Wasserportal: Water level	*Yes, namely, metadata standards and naming conventions*
KWB	Berliner Wasserportal: Water temperature	*Yes, namely, metadata standards and naming conventions*
KWB	KWB: Oxygen concentration	*Yes, namely, metadata standards and naming conventions*
KWB	Sentinel-2	*Yes, namely, use of metadata standards for documentation*
KWB	KWB: Trophic state estimation	*Yes, namely, metadata standards and naming conventions*
Partner	**Dataset**	**What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes, adopting data vocabularies for Land cover maps and connectivity habitats.*
ASTON	Interaction Flow (IF) index	*Vocabularies from GEMET and Essential Variables will be used*
ECMWF	Sensor Community Air Quality data	*Vocabularies from EOINE will be used*
KWB	Crowdwater Data	*Essential Water Variables will be used*

KWB	DWD Weather Data	*Essential Water Variables will be used*
KWB	Berliner Wasserportal: Water level	*Essential Water Variables will be used*
KWB	Berliner Wasserportal: Water temperature	*Essential Water Variables will be used*
KWB	KWB: Oxygen concentration	*to be decided in the project consortium*
KWB	Sentinel-2	*CEOS Analysis ready data vocabularies will be used.*
KWB	KWB: Trophic state estimation	*to be decided in the project consortium*
Partner	**Dataset**	**Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes*
ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Yes*
KWB	Crowdwater Data	*Yes*
KWB	DWD Weather Data	*Yes*
KWB	Berliner Wasserportal: Water level	*Yes*
KWB	Berliner Wasserportal: Water temperature	*to be decided in the project consortium*
KWB	KWB: Oxygen concentration	*to be decided in the project consortium*
KWB	Sentinel-2	*Yes*
KWB	KWB: Trophic state estimation	*to be decided in the project consortium*
Partner	**Dataset**	**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?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes*
ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Tbd*
KWB	Crowdwater Data	*Yes*
KWB	DWD Weather Data	*Yes*
KWB	Berliner Wasserportal: Water level	*Yes*
KWB	Berliner Wasserportal: Water temperature	*to be decided in the project consortium*
KWB	KWB: Oxygen concentration	*to be decided in the project consortium*
KWB	Sentinel-2	*Yes*

KWB	KWB: Trophic state estimation	*to be decided in the project consortium*

3.4 RE-USE OF DATA

Similarly, datasets generated within AD4GD need to be reusable, meeting the following minimum set of criteria[9]:

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 Green Deal strategies.

To enhance reusability, the semantic models that are and 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. 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.

The table below summarises commitment of the partners to ensure data reusability.

Table 17: Summary of partners' commitment to ensure data reusability

Partner	Dataset	How will the data be licensed to permit the widest re-use possible?
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	Yes
ASTON	Interaction Flow (IF) index	*Public license*

[9] Ibid no 4.

ECMWF	Sensor Community Air Quality data	*This data has already a license defined by the producer*
KWB	Crowdwater Data	*Public license as the original data*
KWB	DWD Weather Data	*Public license as the original data*
KWB	Berliner Wasserportal: Water level	*Public license as the original data*
KWB	Berliner Wasserportal: Water temperature	*Public license as the original data*
KWB	KWB: Oxygen concentration	*Public license as the original data*
KWB	Sentinel-2	*Public license as the original data*
KWB	KWB: Trophic state estimation	*Public license as the original data*
Partner	**Dataset**	**Will the data be made available for re-use?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	Yes
ASTON	Interaction Flow (IF) index	*Yes, the data will be made available for re-use*
ECMWF	Sensor Community Air Quality data	*Yes the data will be made available for re-use and there are no plans to request and embargo on its release.*
KWB	Crowdwater Data	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
KWB	DWD Weather Data	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
KWB	Berliner Wasserportal: Water level	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
KWB	Berliner Wasserportal: Water temperature	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
KWB	KWB: Oxygen concentration	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
KWB	Sentinel-2	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Data can directly be re-used as soon as it is uploaded at GeoNetworks*
Partner	**Dataset**	**Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*Yes*
ASTON	Interaction Flow (IF) index	*Yes*
ECMWF	Sensor Community Air Quality data	*Yes*
KWB	Crowdwater Data	*Yes*
KWB	DWD Weather Data	*Yes*

KWB	Berliner Wasserportal: Water level	*Yes*
KWB	Berliner Wasserportal: Water temperature	*Yes*
KWB	KWB: Oxygen concentration	*Yes*
KWB	Sentinel-2	*Yes*
KWB	KWB: Trophic state estimation	*Yes*
Partner	**Dataset**	**How long is it intended for the data to remain re-usable?**
CREAF	Land Use Land Cover Maps from Catalonia (MUCSC)	*For the project dataset: As long as the GeoNetworks exists*
ASTON	Interaction Flow (IF) index	*For the project dataset: As long as the GeoNetworks exists*
ECMWF	Sensor Community Air Quality data	*ECMWF generally keeps indefinite archives of data that it makes available publicly.*
KWB	Crowdwater Data	*For the project dataset: As long as the GeoNetworks exists*
KWB	DWD Weather Data	*For the project dataset: As long as the GeoNetworks exists*
KWB	Berliner Wasserportal: Water level	*For the project dataset: As long as the GeoNetworks exists* *The original data is reusable via the API, however the high temporal resolution data is only accessible for 1 year*
KWB	Berliner Wasserportal: Water temperature	*For the project dataset: As long as the GeoNetworks exists* *The original data is reusable via the API, however the high temporal resolution data is only accessible for 1 year*
KWB	KWB: Oxygen concentration	*For the project dataset: As long as the GeoNetworks exists*
KWB	Sentinel-2	*For the project dataset: As long as the GeoNetworks exists*
KWB	KWB: Trophic state estimation	*For the project dataset: As long as the GeoNetworks exists*

4 GEOSS PRINCIPLES OF DATA MANAGEMENT AND DATA SHARING

In order to best align the project's activities with the Open Science principle, AD4GD also considers the complementary guiding principles provided by the Group on Earth Observations with regards to data management[10] and data sharing[11]. As such, as far as **data management** is concerned, the Consortium designs and will continue to 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 takes into account the above-described principles and incorporates them as much as possible to its activities and architecture, while respecting legal and ethical requirements, as analysed in Section 6. Given the expected evolution of the datasets as the project progresses, further updates might be expected in future iterations of the Data Management Plan in accordance with the development of the regulatory and ethical framework. The above principles are also be balanced against the partners' potential IPR as will be described in Section 7 of the present document.

[10] GEO Data Management Principles Task Force, 'GEOSS Data Management Principles' (28 April 2015).
[11] Data Sharing Working Group of the Group on Earth Observations, 'GEOSS Data Sharing Principles post 2015' (10 March 2014).

5 ETHICAL AND LEGAL ASPECTS

5.1 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.

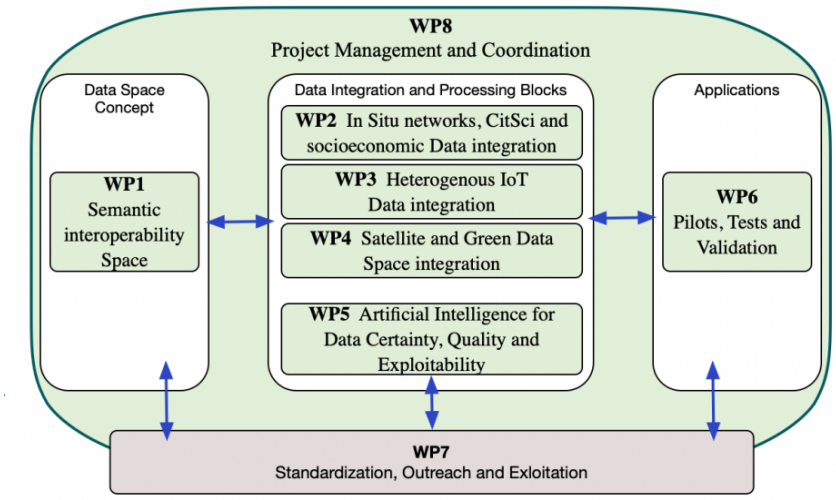

Figure 1: Workplan structure and interactions

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 were asked to provided with **Data Management Questionnaire (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".

5.1.1 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 organisation, 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 identified through the Questionnaires displayed in Annex I;
- **Overviewing the project's compliance** with the GDPR (where personal data processing is involved) 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 analyse 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.

5.2 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 data from non-personal. In order to facilitate this, **an overview of relevant definitions and concepts of the European data protection legal framework is provided in Annex II** of this document.

5.2.1 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 organised, 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 aim at assisting with clarifying partners' role in personal data processing activities carried out in the context of the project.

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 optimisation of biodiversity corridors and the enhancement of greenhouse gas emissions monitoring through dynamic calibration of low-cost sensors. Where personal data is involved, the relevant actors will need to follow the guidelines analysed below. The sections below will further provide an overview of the commitment of the partners (processing personal data) to respect these principles and guidelines.

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. The DPO 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.

5.2.2 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 provides guidance regarding the main obligations when assuming either of these roles.

5.2.2.1 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:

- **Lawfulness, fairness, and transparency** (Article 5(1)(a) GDPR): Controllers must process personal data lawfully, fairly, and in a transparent manner in relation to the data subject
- **Purpose limitation** (Art. 5(1)(b) GDPR): 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 minimisation** (Art. 5(1)(c) GDPR) : 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** (Art. 5(1)(e) GDPR): 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** (Art. 5(2) GDPR): 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 protection by design and by default** (Art. 25 GDPR): 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** (Art. 28 GDPR): 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;
- **Security of data processing** (Art. 32 GDPR): Controllers shall ensure data security, integrity and confidentiality, avoiding to the greater extent possible accidental or unlawful destruction or loss, alteration, unauthorised disclosure of or access to data;
- **Duty of cooperation** (Art. 31 GDPR): Controllers shall cooperate, when requested, with the Data Protection Authority;
- **Data breaches notification** (Art. 33 GDPR): 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.

5.2.2.2 AS JOINT CONTROLLERS

Whenever two or more partner organisations 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** (Art 26 GDPR), 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.

5.2.2.3 AS PROCESSORS

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

- **Confidentiality** (Art. 28(3)(b) GDPR): Ensuring that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality;
- **Sub-processors** (Art. 28(2) and 28(4) GDPR): 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** (Art. 28(3)(g) 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;
- **Security measures** (Art. 32 (1) GDPR): Adopting all the technical and organisational measures to ensure a level of security appropriate to the risk of processing;
- **Record of processing activities** (Art. 30(2) GDPR): Maintaining a record of any processing activities carried out on behalf of the controller;
- **Duty of assistance** (Art. 28(3)(e) and (f) GDPR): Assisting the controller in ensuring compliance with the controller's obligations;
- **Duty of cooperation** (Art. 31 GDPR): Cooperating, when requested, with the DPA.

5.3 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 together with an overview of the commitment of the relevant partners undertaken to respect such principles.

5.3.1 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 European Data Protection Board. In addition to the obligations specified above for data controllers and data processors, a number of further general principles have 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; or
    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 controller's interests.
3. **Transparency**[12]: 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.
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[13].
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 minimise the use of personal data** as much as possible, prioritizing the use of anonymised or aggregated data;

[12] WP29, 'Guidelines on transparency under Regulation 2016/679' (WP260 rev.01, 11 April 2018) [26].
[13] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (version 2.0, 20 October 2020) [35].

3. **To adopt adequate technical and organisational measures,** including anonymisation and/or pseudonymisation 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 recognised certification scheme, where applicable.

5.3.2. DATA SUBJECT RIGHTS

Art. 8 of the Charter of Fundamental Rights of the EU[14] provides the core elements of the right to personal data protection. These core elements 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** (Art. 13 and 14 GDPR), permitting data subjects 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;
b. **Right to access their personal data and obtain a copy** (Art. 15 GDPR) of the information referring to them[15];
c. **Right to rectification** (Art. 16 GDPR) of their personal data where it is inaccurate or incomplete;
d. **Right to object** (Art. 21 GDPR) 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")** (Art. 17 GDPR), giving them the right to erase their data, unless an exception applies;
f. **Right to restrict** the processing of their data (Art. 18 GDPR);
g. **Right to data portability** (Art. 20), where applicable, in a commonly used and machine-readable format[16];
h. **Right to not be subject to automated decision-making and profiling** (Art. 22 GDPR)**,** 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

[14] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
[15] EDPB, 'Guidelines 01/2022on data subject rights - Right of access' (version 1.0, 18 January 2022) [46].
[16] WP29, 'Guidelines on the right to data portability' (WP 242 rev.01, 5 April 2017) 9-10.

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[17]:

- **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.
- **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 **should 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

[17] 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.

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.** Said DPIA will be **maintained and updated throughout the project's lifecycle** in order to ensure that the data processing activities performed remain **compliant**.

There are currently two partners who collect personal data, IoT Lab and Fraunhofer Institute for Applied Information Technology. The table below provides further information on the purposes of the data collection and processing as well as an overview of the safeguards implemented by each of the relevant partners.

Table 18: Overview on personal data collection and processing and implemented safeguards

Partner	**Controller/Processor**
Fraunhofer Institute for Applied Information Technology	Controller
IoT Lab	Controller
Partner	**How will the personal data be collected?**
Fraunhofer Institute for Applied Information Technology	• directly from data subjects who belong to research team • directly from data subjects outside research team (i.e., early adopters, beta testers, etc.) • indirectly through partners of the project Personal data will be collected from subjects inside and outside of the AD4GD research team and through project partners. Collected data will include but is not limited to names, affiliations, positions, and contact information e.g. E-Mail and phone numbers.
IoT Lab	directly from data subjects who belong to your research team
Partner	**Purposes of collecting personal data and relevance to the project's objectives**
Fraunhofer Institute for Applied Information Technology	To satisfy requirements for project related developments personal data of stakeholders needs to be collected to conduct interviews or submit questionnaires. Requirements collection is a detrimental task of the project and foundation of key work packages. Same goes for evaluations of developed deliverables through stakeholder interactions e.g. interviews or user tests. Only through personalized interaction with stakeholders' relevant information can be gathered as basis for research and development tasks in the project.
IoT Lab	Collected personal data are the first name, the last name, the email address and the password of a project partner. The collected personal data are used to create personal accounts on GitLab
Partner	**Further data processing beyond the purposes the data was originally collected for**

Fraunhofer Institute for Applied Information Technology	No further processing beyond the original purposes
IoT Lab	No further processing beyond the original purposes
Partner	**How will data subjects be informed about the purpose of data processing of their personal data in the project?**
Fraunhofer Institute for Applied Information Technology	Individuals are informed within questionnaires through information text. Information sheets are provided ahead of time via E-Mail. At the beginning of interviews individuals are additionally verbally informed about the purpose of the data processing of their personal data.
IoT Lab	By email. Each new user of GitLab must request a new account for GitLab by email with the following required information: first name, last name and email address
Partner	**Consent: collection and documentation**
Fraunhofer Institute for Applied Information Technology	Consent will be collected and documented either in written from when individuals sign provided consent forms prior to taking part in interviews and questionnaires or at the beginning of interviews, that are recorded via MS Teams, consent will be asked for verbally. Informed consent for personal data preservation for the duration of the project is included in the relevant questionnaires: "[...] Contact data will be stored separated from interview data. After the research project ended, contact data will be deleted. Participants have the right to request information about their data collected during the interview at any time. Furthermore, they can request the correction and deletion of such data [...]"
IoT Lab	After the creation of an account on GitLab by the administrator, the new user receives an email which is used to set the user's password. The user can accept or refuse to create his own password.
Partner	**TOMs in place**
Fraunhofer Institute for Applied Information Technology	Within our organisation only a select group of project members has access to the interview documents and therefore the personal data the interview subjects. Furthermore, any information gathered from the interviews for further processing in the projects (e.g. requirements definition) will be anonymised. Confidentiality (Article 32 Paragraph 1 Point b GDPR) Physical Access Control No unauthorised access to Data Processing Facilities: keys, key regulations, entrance security staff (control at the gate, logging of visitors). • Electronic Access Control No unauthorised use of the Data Processing and Data Storage Systems: secure passwords, automatic blocking/locking mechanisms, two-factor authentication, encryption of data carriers/storage media. Use of VPN technology, use of firewalls • Internal Access Control (permissions for user rights of access to and amendment of data) No unauthorised Reading, Copying, Changes or Deletions of Data within the system: rights authorisation concept, need-based rights

	of access. • Isolation Control The isolated Processing of Data, which is collected for differing purposes, e.g. multiple Client support, functional separation between Production / Test; Integrity (Article 32 Paragraph 1 Point b GDPR) • Data Transfer Control No unauthorised Reading, Copying, Changes or Deletions of Data with electronic transfer or transport: Encryption, Virtual Private Networks (VPN), electronic signature; • Data Entry Control Verification, whether and by whom personal data is entered into a Data Processing System, is changed or deleted: Logging, Document Management Availability and Resilience (Article 32 Paragraph 1 Point b GDPR) • Availability Control Prevention of accidental or wilful destruction or loss.: Backup Strategy (online/offline; on-site/off-site), Uninterruptible Power Supply (UPS), virus protection, firewall, reporting procedures and contingency planning • Rapid Recovery (Article 32 Paragraph 1 Point c GDPR) (Article 32 Paragraph 1 Point c GDPR); Procedures for regular testing, assessment and evaluation (Article 32 Paragraph 1 Point d GDPR; Article 25 Paragraph 1 GDPR) • Data Protection Management; • Incident Response Management; • Data Protection by Design and Default (Article 25 Paragraph 2 GDPR); • Order or Contract Control No third party data processing as per Article 28 GDPR without corresponding instructions from the Client: clear and unambiguous contractual arrangements, formalised Order Management, strict controls on the selection of the Service Provider, duty of pre-evaluation.
IoT Lab	The access to the personal data is limited to the administrators of GitLab, who are employees of IoT Lab.
Partner	**Data storage and retention period**
Fraunhofer Institute for Applied Information Technology	The documents containing the personal will be stored in our project's MS Teams folder with restricted access to only current project team members. Personal data will be stored separated from e.g. interview or transcript data. Documents containing personal data will be deleted/destroyed after the project concludes and only aggregate documents (e.g. publications, deliverables) with anonymized information retained.
IoT Lab	The personal data are stored in the GitLab server managed by IoT Lab in Geneva.

	The data will be stored until the user does a request to quit GitLab.
Partner	**Measures to anonymise and/or pseudonymise personal data**
Fraunhofer Institute for Applied Information Technology	Interview subject information will be anonymised when transcripts of the interviews are generated. Any further use of the transcripts and the information therein will no contain personal data. Personal information, such as names, company names, E-Mail addresses or residence, will not be shared beyond the project team. Voluntarily provided contact data such as E-Mail addresses, accounts names etc. will only be used again if the participants declared their will to further participate in the study.
IoT Lab	The passwords are confidential: the administrators cannot see them. More technical details on password encryption used by GitLab are available at https://docs.gitlab.com/ee/security/password_storage.html

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, the main ones of which are as follows:

a. **No personal 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.

Based on the results of the conducted questionnaire, no ethical or legal issues that can have an impact on data sharing have been identified by the partners.

5.5.1. ETHICS GUIDELINES FOR TRUSTWORTHY AI

As already stated, the European Commission has recognised that the use of AI can have a significant impact on the achievement of the goals of the Green Deal[18]. 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

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

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[19] 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.

The European Commission has also made a legislative proposal for the EU Regulation on the Harmonised Rules of AI[20], also known as the AI Act, aiming at enacting a horizontal regulation of Artificial Intelligence. The main objective of this proposed Regulation, is to lay a common legal framework for development, marketing, and the use of AI products and services in the EU.

In accordance with its objectives, in particular Objective 5, AD4GD aims at enhancing certainty, quality, and exploitability of heterogeneous data by leveraging on data analytics, machine learning, and Artificial Intelligence. The project thus plans to employ cutting-edge Deep Learning advancements to extract insights from the shared Data Space for Earth Observation, which will be created through the fusion and harmonisation of heterogenous information originating from diverse sources. To this end, the project aims at exploring and optimising different AI techniques to leverage the capabilities of the computing continuum for Earth system modelling through an existing AI platform covering the complete life cycle of the Machine Learning models as baseline. Keeping this in mind, it is necessary to identify the applicable compliance requirements of AD4GD with the AI Act.

As of January 2024, the proposal is still on the stage of consideration and may undergo changes; some potential obligations that could be relevant to AD4GD can nevertheless be identified at this stage already.

The AI Act lays down a risk-based approach, dividing all the technologies according to the degree of risk: unacceptable risk, high risk, limited risk, and minimal risk. Based on the conducted self-assessment, there might be a number of obligations deriving from the conclusion on the risk level of an AI system in question.

For instance, if it is identified that high-risk AI systems are involved within the project, specific obligations will apply, such as conducting risk assessments, ensuring quality data and human oversight, and registering the systems in a dedicated database. In this regard, appropriate data governance and data management procedures should apply. Similarly, the datasets must be relevant, representative, accurate and complete and, to the extent necessary for the intended purpose, correspond to the characteristics or elements specific to the particular geographic, behavioural or functional context in which the high-risk AI system is intended to be used.

The AI Act additionally lays down transparency obligations on the providers of the AI systems, i.e., AI systems provide transparent and explainable outcomes. In accordance with Article 52 of the current version of the proposal (COM/2021/206 final), providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that

[19] High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (8 April 2019).
[20] 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

they are interacting with an AI system, unless this is obvious from the circumstances and the context of use.

Research partners shall develop and use AI respecting the requirements set, as summarised 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.
- **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[21] 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[22].

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.
- The requirement to **erase or anonymise 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[23] and its limitations[24], 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 (DGA)[25], 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

[21] 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.

[22] 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.

[23] Article 12 of the draft ePrivacy Regulation.

[24] Article 13 of the draft ePrivacy Regulation.

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 Data Act[26] entered into force on 11 January 2024 and shall apply from 12 September 2025[27]. It 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 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.

[25] 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.
[26] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)
[27] *Ibid*. Article 50

The Data Act also introduces **essential requirements regarding interoperability** to be complied with by operators of data spaces, 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.

5.6.5. DATABASE DIRECTIVE

The Database Directive[28] 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 generates 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"[29]. 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 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

In order to establish an interconnected data space, the AD4GD project resorts to open data sets (e.g., GeoNetwork), leveraging the European Open Science Cloud infrastructure and its extensive collection of data. AD4GD's integration of data from various sources, including remote sensing, Virtual Research

[28] 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.
[29] Art. 1 (2) of the Database Directive.

Environments, IoT, and socio-economic data, can potentially involve the reuse of public sector information. Sharing its findings and data to foster the creation of the European Green Data Space is one of the project's primary objectives. Against this backdrop, the **Open Data Directive**[30] **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 in December 2022 **the European Commission adopted a list of high-value datasets**[31] **which should be made available in machine-readable formats and free of charge through APIs.** The datasets are selected from within thematic categories, which include datasets on earth observation and environment. AD4GD will adapt its activities accordingly where required.

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[32] 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**[33] **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 analysing any foreseeable risks, evaluate them and adopt mitigation

[30] 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].

[31] Commission Implementing Regulation of 21.12.2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use, C(2022) 9562 final and the corresponding Annex

[32] 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.

[33] 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."

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 is processed in the context of the AD4GD project. The provisions set by the Regulation on the free flow of non-personal data[34] 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),[35] from 2016, was the first piece of 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 on measures for a high common level of cybersecurity across the Union', from 2022, **NIS2**[36] **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.

[34] 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.
[35] 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.
[36] 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.

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 might 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 brought 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 authorisation;
- **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.

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 are generated in the context of the project are 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 prioritises 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.

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 prioritise its amicable resolution.

7. CONCLUSION AND FUTURE WORK

The current deliverable describes the progress associated with the data already involved or that is anticipated to be involved in the course of the project in its current stage of development, as well as the steps to ensure compliance with the Open Science and FAIR principles, ethical and legal provisions. This deliverable also provides further guidance on the obligations arising from the relevant legal and ethical framework in order to facilitate partners when finalising their data management strategy and points of action.

The information reported in the present deliverable is expected to be further expanded and complemented by the further iteration of the Data Management Plan when more information on the partner's precise workplan is available. It is also expected that the datasets already produced will evolve, therefore the next iteration of the DMP will also address these updates. The Questionnaire provided in Annex I of the present deliverable will serve as the baseline for the partner's work and the collection of the progress and updates on the already provided answers. The relevant updates and elements will be reported in the third 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.

Commission Implementing Regulation (EU) of 21.12.2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use, C(2022) 9562 final

ANNEX I – DATA MANAGEMENT QUESTIONNAIRE

PART A – DATASET DESCRIPTION

	Please provide your answers in this column:
Name of the used dataset(s)	*Click or tap here to fill in*
Reference to GeoNetwork record	*Click or tap here to fill in*
Short description of the dataset(s)	*Click or tap here to fill in*
Purpose for which you use/ process the dataset(s)	*Click or tap here to fill in*
Format(s) of dataset(s)	*Click or tap here to fill in*
Where will you store the dataset(s)?	*Click or tap here to fill in*
What is the main source of the dataset(s)?	*Click or tap here to fill in*
Who owns the dataset(s)?	*Click or tap here to fill in*
Origin of the dataset	*Click or tap here to fill in*
Are there any restrictions for the use of the datasets?	*Click or tap here to fill in*
Who has access to the datasets?	*Click or tap here to fill in*
Under which licence did you obtain access to the datasets?	*Click or tap here to fill in*
Additional comments	*Click or tap here to fill in*

PART B – PERSONAL DATA

1. **Indicate the type of dataset you intend to manage (create or process):**

☐ **non personal data**
(data that does not contain any information that can be used to identify a natural person, e.g., environmental data)

☐ **personal data**
(any information relating to identified or identifiable individuals, including e.g., email or IP addresses)

Please, specify

☐ **special categories of data**
(*personal data revealing sensitive information such as sexual orientation, racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, as well as any health, genetic or biometric data related to the data subjects)*

Please, specify

2. **If applicable, how will you collect these personal data?**

☐ directly from data subjects who belong to your research team
☐ directly from data subjects outside your research team (i.e., early adopters, beta testers, etc.)
☐ indirectly through partners of the project
☐ indirectly through other organisations external to the project
☐Other
Please, specify

3. **If personal data is collected, describe the purpose(s) of collecting personal data and the relevance to the objectives of the project.**

4. **Will you process the generated data for any further purposes than the ones it was originally collected for?**

☐ Yes
☐ No

If yes, please describe the purpose of this additional processing:

5. **How will you inform the individuals (the data subjects) about the purpose of the data processing of their personal data in the project?**

6. **How do you plan to collect and document the consent of the data subjects whose personal data will be processed by you?**

Please, note that consent must be:

- *freely given,*
- *specific,*
- *informed,*
- *unambiguous,*
- *and can be revoked.*

7. **What technical and organisational measures have you put in place to ensure personal data security?**

(i.e., anonymisation techniques, pseudonymisation, tokenisation, etc)

8. **How and where will you store the personal data?**
9. **Is the data safely stored in certified repositories? Please, describe.**
10. **For how long do you intend to keep personal data?**
11. **What technical and organisational measures (TOMs) are in place to protect and secure the personal data?**

(e.g., pseudonymisation, encryption, data protection impact assessments, data protection and security policies and procedures, etc.)

12. **Describe the measures in place to anonymise and/or pseudonymise the personal data whenever possible?**

13. **Do you plan to:**

- to share personal data with other partners inside the project?

☐ Yes
☐ No

If you answered yes, please, specify: > to whom, > for what purpose and task, > whether the data will be anonymised/psedonymised or raw, > and whether you have an agreement in place with the respective party.

- share personal data with data processors (third parties who will process data under your control)?

☐ Yes
☐ No

If you answered yes, please, specify: > to whom, > for what purpose and task, > whether the data will be anonymised/psedonymised or raw, > and whether you have an agreement in place with the respective party.

- transfer personal data to countries outside of Europe?

☐ Yes
☐ No

If you answered yes, please, specify the country and the purpose.

- make personal data available to third parties for further research or processing?

☐ Yes
☐ No

If you answered yes, please, specify: > to whom, > for what purpose and task, > whether the data will be anonymised/psedonymised or raw, > and whether you have an agreement in place with the respective party.

- perform a data protection impact assessment?

☐ Yes
☐ No

14. **Are you in charge of making decisions about what data to be collected/ processed, how and for what purpose?**

☐ Yes, namely,

☐ No

15. **Do you process data under another partner's behalf/instructions?**

☐ Yes, namely,

☐ No

16. **Will you reuse data or existing datasets for further purposes beyond the project?**

☐ Yes

☐ No

PART C – FAIR DATA

The FAIR principles stand for Findable, Accessible, Interoperable, and Reusable, and provide a framework for improving the management and usability of research data. These principles aim to enhance the value and impact of data by making it more discoverable, accessible, and effectively reusable for both humans and machines.

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 yes, please describe.**

☐ Yes

☐ No

If you answered yes, please, describe.

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

3. **Will search keywords be provided that optimise possibilities for re-use?**

☐ Yes

☐ No

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

☐ Yes

☐ No

5. **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?**

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

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

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

6. **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.

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

8. **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?**

2. **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.**

3. **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

Further comments:

4. **How long is it intended for the data to remain re-usable?**

PART D – ALLOCATION OF RESOURCES

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

☐ Yes
☐ No

If you answered yes, please, specify

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

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

PART E – ETHICAL & LEGAL ASPECTS

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

☐ Yes
☐ No

If you answered yes, please, describe

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

☐ Yes
☐ No

If you answered yes, please, describe

PART F – INTELLECTUAL PROPERTY RIGHTS (IPR) OF THE DATASET BROUGHT INTO THE PROJECT BY A PARTNER OR PRODUCED IN THE PROJECT

Please, explain the Intellectual Property Rights (IPR) brought to the project or generated through it (this includes patents, trademarks, software licensing, copyright, etc)

1. **Is the dataset subjected to intellectual property (IP) in the AD4GD project?**

☐ Yes
☐ No

If you answered yes, please, describe. Answer no if the dataset is not brought into the project by a partner or produced in the project.

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

3. If the dataset is subjected to IP-protection-eligible, how will you exploit the dataset?

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

☐ Yes
☐ No

If you answered yes, please, specify for how long

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 relevant to 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. Anonymised data as well as data relating to legal persons are not considered 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 AGREEMENT	Art. 26 GDPR requires joint controllers to adopt an arrangement 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 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.