

D1.1 Data Management Plan

Actual Submission Date: **11/25/2024**

Produced by: **SIMAVI**

Accurate

<https://accurateproject.eu/>

HORIZON 1.0– 2023-12-04

ACCURATE — HORIZON-CL4-2023-TWIN-TRANSITION-01

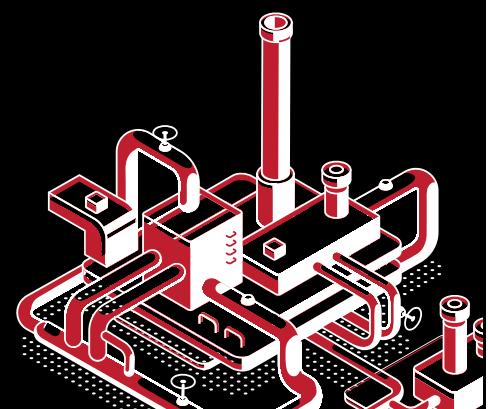
Grant Agreement no.: 101138269

Start date of project: 01/12/2023 - Duration: 36 months



**Funded by
the European Union**

The ACCURATE project is funded by the European Union, under Grant Agreement number 101138269. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



DELIVERABLE FACTSHEET

Deliverable D5.1	
Nature of the Deliverable:	Public
Due date of the Deliverable:	M6 – 31/07/2024
Actual Submission Date:	M10 – 14/10/2024
Produced by:	SIMAVI
Contributors:	N/A
Work Package Leader Responsible:	SIMAVI
Reviewed by:	Internal SIMAVI

Dissemination level	
	PU = Public
PU	PP = Restricted to other programme participants (including the EC)
	RE = Restricted to a group of the consortium (including the EC)
	CO = Confidential, only members of the consortium (including the EC)

Contents

Terms and abbreviations	5
Public Summary	5
1 Introduction	5
1.1 About this Deliverable	5
1.2 Document Structure	6
2 Data Summary	7
2.1 Data related to the pilot cases	7
2.2 Data related to scientific publications	7
2.3 Accurate public deliverables	8
3 Acces to Data	9
3.1 Data related to the pilots cases	9
3.2 Data related to scientific publication	9
3.3 Data related to deliverables	9
4 Resources, Security and Ethical Aspects	10
4.1 Allocation of resources	10
4.2 Data Security and Protection	10
4.3 Ethical aspects	10
5 Accurate data set description	10
6 Archiving Data and Preserving Infrastructure	12
6.1 Sharepoint repository	12
6.2 Project website	12
6.3 Code Repository	13
7 Conclusion	13

ACCURATE

Figures

Terms and abbreviations

DMP	Data Management Plan
DoA	Description of Action
EC	European Commission
EU	European Union
FAIR	findable, accessible, interoperable and reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
ORD	Open Research Data
RRI	Responsible Research and Innovation

Public Summary

This deliverable provides the initial version of the plan for managing, collecting, generating, storing, and preserving data related to Accurate Project activities. The action encompasses three types of data: data related to pilot cases, data from publications, and data from public deliverables.

Any specific closed or proprietary data will be considered, addressing issues related to openness. While there is currently no such data, it will be accounted for in the future, along with its implications for IPR management and the approach to exploitable results.

The document follows the EC template and outlines how these various types of data will be collected, identifies the primary beneficiaries, and describes how the Accurate Project will store and manage the data. It also addresses whether the data will be made accessible, findable, and reusable. Additionally, the text discusses the resources required to ensure both the openness and security of the data within the context of Accurate.

This document serves as the first version of the data management plan and will be updated in accordance with the project's progress, including technological developments.

1 Introduction

1.1 About this Deliverable

This deliverable addresses data management in the Accurate Project. Within this project, we have identified four distinct types of data: (1) data related to publications, (2) data from public deliverables produced during the research activities, (3) data collected from pilot cases, and (4) general project advancements, including technological choices, which will be utilized in implementing the various project outcomes. This deliverable is complementary to deliverable D1.4 – Ethics Report – which will be delivered in M36.

ACCURATE

The context of this deliverable is summarized in the bellow table:

Project Item	Description
Objectives	The main objective of this deliverable is to provide a first version of the description for the data management life cycle
Exploitable results	N/A
Work plan	This deliverable is the main result from T1.3. It defines and addresses the data management plan (DMP) of the project.
Impacted packages	Transversal all WP's except WP8
Deliverables	<p>D2.1 Requirements specification and ontology service for semantic representation of components</p> <p>D3.1 Human-Centric Decision Support System for MAAS Production Adaptation</p> <p>D 4.1 Simulation and optimization model for supply chain management and stress test</p> <p>D5.2 Compute-to-Data Environment WP5 9 - DAO DEM — Demonstrator, pilot, prototype</p> <p>D5.3 Sovereign Data Sharing</p> <p>D6.4 ACCURATE framework - initial version</p> <p>D7.1 Pilots 'deployment strategy</p> <p>D7.2 Validation report</p>

1.2 Document Structure

The document adheres to the H2020 template for a Data Management Plan (DMP) . Section 2 summarizes the objectives of data collection and generation for ACCURATE. Section 3 details how the data and metadata will be made FAIR—ensuring they are accessible, findable, and reusable. Section 5 includes a template for the anticipated datasets for Accurate Project. Section 6 offers conclusions and outlines future work.

2 Data Summary

At its core, the Accurate project will gather and analyze data from various sources.

Further information regarding data usage, best practices, ethical principles, and fundamental rights outlined in the European Union's data protection regulations, directives, and opinions is detailed in the project's Grant Agreement (sections 5.1 and 5.1.1).

In the following sections, this deliverable addresses specific aspects related to data from the pilot cases, scientific publications, and public deliverables.

2.1 Data related to the pilot cases

This section clarifies the significance of collecting and generating data within the ACCURATE pilot cases, identifying the relevant data sources and their importance for achieving the project's objectives.

The initial, non-exhaustive list of data types and categories collected from the pilot cases includes:

- **Personal Data:** This is generated through interactions with ACCURATE and can originate from various sources, including:
 - **Obtrusive Data:** This type requires direct interaction from individuals, making them aware that they are being studied, which may influence their responses or behavior. Examples include:
 - Questionnaires and interviews for data collection
 - Wearable sensors
 - Other IoT devices and sensors
 - Interactions with the ACCURATE app and dashboard
 - **Unobtrusive Data:** This is generated indirectly, requiring no explicit interaction from individuals, meaning they are unaware of being studied. As a result, their responses and behavior remain unaffected. An example includes presence or location sensors.
- **Environmental Data:** This data is collected without human intervention, such as readings from factory environment sensors (e.g., humidity, temperature, air pollution).
- **Public Data:** This includes data already available in public repositories, such as Smart City data (e.g., traffic, transportation, public events) and Smart Environment data (e.g., real-time pollution levels, forest fire detection, weather forecasts).

2.2 Data related to scientific publications

ACCURATE will publish scientific publications in conferences and journals as part of the planned dissemination activities.

Scientific publications' data are made available often using accessible PDF files. The metadata to be used will be compliant with the format requested by OpenAIRE as well as the one requested by the repository where the papers are to be deposited.

2.3 Accurate public deliverables

All information and material related to the public, such as **public** deliverables, brochures, posters and so on will be freely available on the project website in the form of accessible PDF files. When IPR of foreground knowledge needs to be protected, the corresponding disclosures will be published.

3 Open Data and FAIR data

Open data is typically defined as research data that is freely accessible, allowing users to access, utilize, reproduce, and share it without charge. According to the European Commission guidelines, projects funded under Horizon Europe, such as Accurate, must guarantee open access to all peer-reviewed scientific publications resulting from their research and manage the generated data responsibly, adhering to the FAIR principles.

FAIR principles—Findable, Accessible, Interoperable, and Reusable—are a set of guiding principles rather than strict standards. This means that all partners in the Accurate project should ensure their research data is organized to be easily findable, accessible, interoperable with other data, and reusable, as outlined in Table below.

F	<u>Easily discoverable through metadata, identifiable, and locatable using a standard identification method.</u>
A	<u>Always accessible and obtainable; even if the data is restricted, the metadata remains open.</u>
I	<u>Both syntactically parseable and semantically clear, facilitating data exchange and reuse among researchers, institutions, organizations, or countries.</u>
R	<u>Well-described and shared under minimally restrictive licenses, enabling the widest possible reuse and facilitating seamless integration with other data sources.</u>

It's essential to mention that adhering to the FAIR principles does not necessarily require data to be openly shared with everyone. For example, data can align with the FAIR principles while remaining private or shared only under specific conditions, meaning they can be FAIR but not openly accessible. Conversely, open data might not be considered FAIR if it is publicly available but lacks proper documentation, such as clear licensing for reuse, which is necessary to meet the FAIR criteria.

4 Acces to Data

In the following section the focused will be on the topic related to the feasibility of making data findable, reusable in the context of Accurate Project

4.1 Data related to the pilots cases

The personal data used in the project will be anonymized, if the case and encrypted so that the entities cannot be revealed. The directive that will be used is Regulation (European Union) 2016/679, and all personal data must be treated as sensitive. In this case, these data will be pseudonymized and in the platform only some code will be displayed.

The flow of data, how the data will be collected and used inside of the project will be presented in some deliverables.

At this stage, all data—both collected and generated—for pilot cases will be pseudo-anonymized (or anonymized if necessary) and will not be made openly accessible. All collected, stored, and processed data will be treated as strictly confidential and retained only for a specified period, as outlined in the consent form. This duration will be limited to what is necessary to achieve the project's objectives and validate the scenario. After this period, any sensitive or personal data will be destroyed as required.

4.2 Data related to scientific publication

For scientific publications, a persistent identification number will be assigned when the publications are uploaded to the chosen repository or repositories

4.3 Data related to deliverables

For the project's publications on the website, the naming convention will be as follows: <<Dx.y Deliverable Name _ Submission Date (YYYYMMDD).pdf>>.

Deliverables will be stored in SIMAVI's sharepoint, for three years beyond the project's duration. Public deliverables will be accessible through the project website once they have been approved by the European Commission.

5 Resources, Security and Ethical Aspects

5.1 Allocation of resources

Accurate does not anticipate any additional resource needs beyond the duration of the project to manage data or to ensure the data is FAIR. As mentioned earlier, preference will be given to open access repositories.

5.2 Data Security and Protection

Accurate Project is committed to ensuring compliance with the General Data Protection Regulation (GDPR), which came into effect in May 2018, particularly regarding the protection of private users' data and adherence to relevant national legislation for each pilot.

The project will follow international data protection laws and good data management practices. Organizations that process data, referred to as 'Data Controllers,' will be obligated to handle data fairly, securely, and for explicit and legitimate purposes. GDPR, particularly Chapter 3 (Art. 12 and onwards), along with national regulations, establishes a set of rights for data subjects. These rights include the right to be informed about the processing of their personal data, the right to access their data, and, when necessary, the right to amend or delete it.

5.3 Ethical aspects

The foundation of ethical research lies in the principle of informed consent. Data will be pseudo-anonymized (or anonymized when necessary) and reported as aggregated data (where applicable) in the documents assessing the outcomes of ACCURATE Participants will have the option to withdraw from the use cases at any time, and their data—including anonymized data—will be deleted.

6 Accurate data set description

This section provides a template for describing all anticipated Accurate datasets, utilizing the Data Management Plan (DMP) template established by the European Commission for Horizon 2020. The definitions of the various dataset categories (see Table 2) underscore the significance of long-term data preservation and the need for the broadest possible sharing of knowledge generated by EU projects.

Questionnaire regarding the collection, access and processing data within the project implementation

<u>Dataset description</u>	<u>Short description of the dataset</u>
<u>Type of data</u>	<u>What type of data will your WP collect, process and keep?</u>

ACCURATE

	<u>Any personal or sensitive data?</u>
<u>Collection of data</u>	<p><u>How will this data be collected? E.g. Directly from researchers, pilot testers, trials, indirectly from the use of technical systems (logins, cookies)</u></p> <p><u>What research methods are you using to collect the data? (e.g. observations, interviews, survey, WorkShop, Delphi method, Vignettes, participation in simulations, etc.)</u></p>
<u>Data format</u>	<u>What format(s) will the data be in?</u>
<u>Data size</u>	<u>Size of the dataset</u>
<u>Ownership</u>	<u>Who owns the data?</u>
<u>Storage</u>	<p><u>Who will store the data?</u></p> <p><u>Where will the dataset be stored?</u></p> <p><u>How will you ensure it is stored securely?</u></p>
<u>Retention period</u>	<u>How long will you keep the data?</u>
<u>Work package</u>	<u>For which WPs will this dataset be used?</u>
<u>Purpose</u>	<u>What will be the purpose of the use of the dataset</u>
<u>Recruiting</u>	<u>How will you identify and recruit people to participate in WP activities?</u>
<u>Processing</u>	<p><u>How will be the data processed and analysed both during and after the completion of the project? (e.g. entering, digitizing, transcribing and translating data, checking, validating, cleaning or anonymizing data, deriving data, describing and documenting data, managing and storing data, analysing and interpreting data, producing research outputs and citing data, etc.)</u></p>
<u>Profiling</u>	<p><u>In case the research involves profiling, provide explanation how the data subjects will be informed of the existence of the profiling, its possible consequences and how their fundamental rights will be safeguarded.</u></p>
<u>Access</u>	<u>Can this data be shared with other project partners?</u>
<u>Availability after-project</u>	<p><u>Can this data be made openly accessible after the close of the project?</u></p> <p><u>If YES, how it will be shared? (e.g. discovery metadada, selecting appropriate access to data, publishing data or promoting data - where?-. If NO, why not?</u></p>
<u>Accessibility and re-use</u>	<p><u>How will be made accessible for verification or re-use? (e.g. conducting secondary analysis, undertaking follow-up research, conducting research reviews, scrutinizing findings, and using data for teaching and learning, etc.)</u></p> <p><u>Are there any standard formats for managing or disseminating the datasets?</u></p> <p><u>How will metadata be generated and captured?</u></p>
<u>Comments</u>	<u>Any other comment</u>

The complete input from the pilots will be presented in the next version of the document.

7 Archiving Data and Preserving Infrastructure

The following section provides brief descriptions of the platforms and repositories selected for ACCURATE data storage and dissemination.

7.1 Sharepoint repository

The SharePoint repository is a collaborative platform used for document management and storage, facilitating efficient data sharing and team collaboration. Key features that will be used in the project:

- **Document Storage and Organization:** Users can create sites and libraries to organize documents in a structured manner, allowing for easy access and retrieval.
- **Version Control:** SharePoint tracks document versions, enabling users to view and restore previous iterations, which is essential for maintaining data integrity.
- **Access Control:** The platform offers granular permissions settings, allowing administrators to control who can view, edit, or share specific documents.
- **Collaboration Tools:** Integrated tools such as discussion boards, task lists, and calendars enhance team collaboration and project management.
- **Search Functionality:** A robust search feature enables users to quickly find documents and information within the repository.
- **Integration Capabilities:** SharePoint seamlessly integrates with other Microsoft Office applications and third-party tools, facilitating a streamlined workflow.
- **Backup and Recovery:** The platform includes automated backup options and disaster recovery features to ensure data security and availability.
- **Compliance and Security:** SharePoint adheres to various compliance standards, ensuring that data storage and sharing practices meet regulatory requirements.

This repository is selected for its alignment with the FAIR data principles, promoting findability, accessibility, interoperability, and reusability of data.

The platform is hosted by SIMAVI.

7.2 Project website

The ACCURATE website serves as the primary public information hub for the project, accessible at <https://accurateproject.eu/>. It features static content, including an overview of the project's concept, goals, and proposed approach, alongside dynamic content such as event announcements, news updates, and blog posts.

A dedicated section, "Published Materials," will be developed and all materials will be visible in this section

ACCURATE

All public information on the ACCURATE website is accessible without restrictions, allowing any visitor to view it without needing to create an account or provide personal data. Additionally, all website-related data is regularly backed up.

7.3 Code Repository

Accurate software components will be categorized into Open Source and Closed Source components.

To achieve this, all Open Source components will be deposited in a public Git-based web repository, making them available to the community for use and enhancement. Closed Source components may be stored in private repositories. Git, as a distributed version control system, fosters sharing and collaboration among users while providing version control for creators. The two main Git-based hosting providers under consideration are GitLab and GitHub.

GitLab is a fully integrated software development platform that offers a range of essential tools for software development, deployment, and project management, including:

- **Code hosting** with version control.
- **Issue Tracker** for new implementations, bug reports, and feedback.
- **Issue Boards** for organizing and prioritizing tasks.
- **Code review** through Merge Requests, featuring live previews of changes per branch with Review Apps.
- **Built-in Continuous Integration (CI), Continuous Deployment (CD), and Continuous Delivery (CD)** support, allowing for automated building, testing, and deployment of applications. Each pull or push triggers a CI pipeline, executing jobs in stages. Jobs within a stage run in parallel, and if they all succeed, the pipeline progresses to the next stage. The typical pipeline consists of four stages: build, test, staging, and production. The status of current and historical pipelines can be monitored in a dedicated Pipelines tab.

In addition to GitLab, the consortium is also considering **GitHub** for collaborative software development of Open Source components. GitHub, a leading Git-based online repository hosting provider, supports various formats and features beyond code storage.

Both GitLab and GitHub provide options for private repositories. GitLab offers unlimited public and private repositories with no transfer limits and allows for unlimited collaborators. GitHub offers free plans for open-source projects and paid plans for unlimited private repositories.

8 Conclusion

This deliverable presents the initial version of the data management plan for the ACCURATE project. It includes a template for describing the datasets that project partners intend to collect and use.

ACCURATE

This document is a living document that will be updated annually throughout the project's duration, with a final version due in M36 as part of the technical reports. It is expected that this final version will clarify aspects that are currently uncertain as work progresses across all work packages.