Data Management Plan

Deliverable Number	D1.2
Deliverable Type	R – Document, Report
Dissemination Level	PU (Public)
Deliverable Responsible	Pius Sonnberger
Planned Delivery Date	2024-06-30 (Month 6)
Actual Delivery Date	2024-06-26 (Month 6)
Document Version & Status	V1.2 | Final

Project Acronym	GIANTS
Project Title	Green Intelligent Affordable New Transport Solutions
Grant Agreement Number	101138220
Project Coordinator	Virtual Vehicle Research GmbH
Project Website	www.giants-project.eu

Funded by the European Union

GIANTS has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101138220.

Authors

Name	Organisation	Name	Organisation
Pius Sonnberger	ViF	Medina Ćustić	ViF

Reviewers

Name	Organisation	Date
Karl Lambauer	HiWiTronics	2024-06-13
Mohamed El Baghdadi	VUB	2024-06-14

Change History

Version	Date	Name/Organisation	Description
V0.1	2024-05-15	Medina Ćustić (ViF)	Draft created; first outline of chapters
V0.2	2024-06-07	Pius Sonnberger (ViF)	Content added; Descripting of used DM tools
V1.0	2024-06-10	Pius Sonnberger (ViF)	References and executive summary added
V1.1	2024-06-25	Pius Sonnberger (ViF)	Review comments added
V1.2	2024-06-26	Medina Ćustić (ViF)	Formal Review and Deliverable Release

Table of Contents

List of Tables

1. Executive Summary

This document is Deliverable 1.2 of the GIANTS (Green Intelligent Affordable New Transport Solutions) project, which contains the initial Data Management Plan. This is a living document and will be delivered in an updated version with D1.3, the Data Management Plan Update in M16 of the project.

GIANTS handles data management in strict accordance with applicable EU and national data protection laws. All project partners must guarantee compliance with the EU's General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) [1] and relevant national regulations.

GIANTS is a multi-disciplinary project with data from various sources, reaching from documents, survey results, simulation-, measurement- and sensor data. This makes a well-structured data management strategy a necessity. The data management plan provides the rulebook how data is treated within the project to make it findable and accessible to achieve the project objectives. All sensitive data that is collected will be handled according to the General Data Protection Regulation (GDPR) [1]. Further, the collected data in the GIANTS project will be compliant with the FAIR data principles.

Keywords: Data Management Plan, FAIR Data, Data handling, Data security

2. Objectives

Data management is closely related to all objectives of the GIANTS project. The overall process of data management is carried out in the Overall Project Management (Task 1.1) of WP1 and is also closely linked to the WP Management (Task 1.2), where the information transfer between the tasks and work packages is handled. The guidelines of how data is handled within the GIANTS project is agreed upon in the Grant Agreement (GA) [2] and Consortium Agreement (CA) [3]. The objective of the data management plan is to give guidelines for the data handling and describe how the data management is performed in the GIANTS project.

Well-structured data management is particularly important for the measured data that is carried out in the demonstration work package (WP6) which feed back to the digital twins that are developed in WP4. This will enable the GIANTS project to achieve the overall objective 4 (Digitalization of components, systems and platform for flexible design and efficient operation of frugal vehicle solutions). The data management tool (Data Context Hub [4]) that is used for the handling of the measured data is described in detail in Chapter 5.

The FAIR data principles will be used by all partners and the data management process will be monitored by the project coordinator (VIF). No personal data will be processed, except for what is necessary for project administration (e.g., contact details). The survey data that is collected in WP2 will be fully randomised and no personal data will be stored or processed.

The updated Data Management Plan (D1.3) is due in M16 of the project and will finalise the overall Data Management Strategy.

3. Data Summary

3.1 Project Deliverables

Deliverables (public and sensitive) will be made available as PDF file (approximately 500 MB) following the Description of Action [2]. Once the deliverable has been accepted by CINEA, the public deliverable will be made available for download via (i) CORDIS [5] and (ii) the GIANTS website [6]. It will also be uploaded to the GIANTS Zenodo [7] community. Sensitive deliverables will not be shared with the public. However, in case the consortium agrees, the executive summary can be made available on the website. The respective data will be generated during the project.

3.2 Promotional Material

The promotional material contains public material to promote GIANTS (e.g. flyers, roll-up, public presentations). This will be made available on the GIANTS website in PDF format (approximately 200 MB). Additionally, it will be stored on the GIANTS SharePoint for the partners to access it there as well. The respective data will be generated during the project.

3.3 Dissemination and Communication Activities, including Publications

For D&C activities, an Excel list has been created to track the planned and performed activities. This list is for internal use only and is available on the SharePoint. Project partners can access and update the list at any time. Additionally, all publications should be open access (gold or green) according to the GA [2] and CA [3] of GIANTS. The publications are tracked in a SharePoint list and is regularly updated. This list can be exported as xml file (approximately 5 MB, depending on the number of items). Publications will receive a DOI (Digital Object Identifier). Additionally, publications will be included in the GIANTS Zenodo community to ensure that the publications can be found quickly. The respective data will be generated during the project.

3.4 Survey Results

A survey will be conducted to understand the users' and stakeholders' needs in work package 2 of the GIANTS project. The retrieved results are processed in Excel spreadsheets and will be stored in .xlsx format (approximately 100 MB). The retrieved results will be stored on the GIANTS SharePoint. Personal information about the participants will not be part of the study data, therefore an identification of the participants from the data is not possible. The outcome of this survey will complement data from already existing, published data from surveys that will be re-used.

3.5 Simulation Data

Digital Twins and simulation models will be built up in work package 4 of the GIANTS project as described in the Grant Agreement [2]. The used simulation and modelling software will be chosen according to the individual problem to be solved. The co-simulation platform ICOS will be used to enable the interdisciplinary use of data. The simulation data (approximately 200GB, depending on the needed simulations) will be described in the corresponding deliverables and publications.

3.6 Measurement Data

In the GIANTS project, one focus point is the demonstration of the developed frugal vehicle solutions. It is foreseen to carry out testing on a component level where data from different sensor systems will be collected and uploaded via a secure connection to a server. Further, the developed solutions will be brought to application at demonstration sites in Bruges, Stockholm, Delhi, Kisumu and Manila. This data will mainly be time series data from different measurement sources leading to an expected data size of 1TB. To process this data from multiple sources, a tool is needed to make the data handleable. Therefore, the integrated platform Data Context Hub (DCH) [4] will be used to streamline the data. A detailed description of this platform is provided in Chapter 5. The collected data will be stored ether on a Kubernetes [8] server hosted at Virtual Vehicle (Coordinator) or on a cloud-based solution (Microsoft Azure [9]). Which of these two solutions will be used, is decided in course of the project and will be described in the updated Data Management Plan (D1.3). All measured vehicle data will be stored on central data logger devices inside the demo vehicles and will be uploaded to the chosen server or cloud-based solution in Europe. This will make sure, that the data that is measured outside of Europe (Delhi, Kisumu and Manila) will be treated according to EU regulations.

4. FAIR Data

4.1 Making Data Findable, including Provisions for Metadata

Scientific publications will be identified with a DOI and made available via ZENODO [7].

In D1.1 “Project Management Handbook” [10], the naming conventions for general documents as well as deliverables have been defined. These guidelines are included in Table 4-1, with further explanations below (taken from D1.1).

Table 4-1: Naming Conventions for GIANTS Documents

Document	Naming Convention and Example
General Documents (reports, presentations…)	yyyymmdd_GIANTS_Title_ _vX.X.doc/pdf/xls (...) 20240201_GIANTS_MoM_KOM_final.pdf
Deliverables	GIANTS_DX.X_Title_vX.X.doc/pdf/xls (...) GIANTS_D1.1_ProjectHandbook_v1.0.pdf

- Title: Short document description (please do not use special characters, e.g. “/” or “:”)
- Date: Date of creation (format: yyyymmdd)
- Version: vX.Y (X = a major release; Y = a minor release; e.g. v0.1, v0.2, v1.0, v2.0 …)
- File Extension: According to the type of the file (docx, pdf, …)

Simulation data will be handled with the ICOS tool. ICOS will create simulation data, based on subsystems provided by different partners, the simulation data will be accessible at the host machine where ICOS/Model.CONNECT is running. The data will be stored based on the defined naming and will be time stamped. By using this meaningful metadata information, Windows Explorer can be used for searching the simulation data. Only the data relevant for the coupling can be seen at the co-simulation level, any extra data must be made available by the subsystem provided through extra outputs.

The Data Context Hub (DCH) platform will be used to make the measured data findable using its search and navigational capabilities.

4.2 Making Data Accessible

A SharePoint provided by the coordinator (ViF) has been set up for the proposal and is being used also for the project duration (and beyond for archiving, see chapter 6). Among other things, it is used to store and access data. The access to the SharePoint, which is managed by ViF, is limited to the project partners. The SharePoint has been chosen in order to monitor the version history, ensure regular backups as well as to restore and manage the data.

Data that is marked as “public” will be made publicly available via the GIANTS website and other channels such as LinkedIn. These materials can be accessed by everyone without requesting access to it. Where possible, data will be made available via data platforms (e.g. ZENODO [7]). Additionally, partners must ensure open access to their publications.

In ICOS, data will be automatically accessible for the partner who is hosting the co-simulation.

The accessibility of the measured data will be provided by DCH, by providing access from different Application Programming Interfaces (API) endpoints.

4.3 Making Data Interoperable

Interoperability in GIANTS is achieved through implementation of standardised naming including metadata information of data files as described before. For publications, only machine-readable formats are used, meaning that scanned files will not be made public. Only files in common data formats .docx, .pdf, .txt, .xlsx etc. will be used for publication purposes.

Moreover, the data collection and aggregation platform Data Context Hub is used to make the sensor data interoperable and usable for multiple purposes. Deeper explanation on the Data Context Hub framework is provided in Chapter 5.

4.4 Increase Data Re-Use

The modelling data handled with ICOS will contain a clear documentation of what is co-simulated, including documentation of the inputs and outputs, whereas this is a highly cooperative task where all the subsystem providers need to provide the appropriate information.

The measured data handled with DCH will be reusable in many tools and environments for all project partners that have access via a test license during the runtime of the project.

Moreover, selected data will be published under open access making it accessible after the project runtime. The confidentiality of the published data will be first coordinated with all relevant project partners.

5. Other Research Outputs

Data Context Hub [4] is the software that will be used as a data collection and aggregation platform for the measurement data in the project. It will provide the features to simplify the data management, which will also conform to FAIR principles: it will make the data and its metadata **findable** by providing search and navigation capabilities, provides a good **accessibility** to the data by different API endpoints and will ensure the **interoperability** between different tools that want to access the data, and finally all these capabilities allow the data to be **reuseable** in many tools and environments.

To fulfil these objectives, DCH will perform the following three steps, which correspond to different stages of data management:

1. **Intake:** Retrieving information from various sources, which may be scattered to different systems, either by data pumps that pull data from a source and pump to the DCH or by a push API that helps to push data directly to the DCH. Both methods can be automated as workflows. The Push API can also be used to retrieve data from streams.
2. **Contextualisation:** Defining domain-specific data models and transformations, which enable the extraction of an explorable knowledge graph. It is also possible to implement additional methods, such as semantic-based or ML-based approaches, for reasoning to infer new facts or insights over the knowledge graph.
3. **Delivery:** Finally, the aggregated data is stored as a knowledge graph and can be used for separate applications for different purposes through DCH APIs. Hence, it is possible to use the data navigation tools from DCH or integrate the context-based navigation into the existing applications over DCH APIs.

The final step will deliver the data to various applications to perform contextual data analysis and reasoning using different methods to enrich the knowledge graph with new facts. It is always possible to use these steps iteratively.

DCH will also provide the means for *contextual data analysis*, which involves the examination and interpretation of data within its specific context or environment to derive meaningful insights, understand relationships, and make informed decisions. Context refers to the circumstances, conditions, or surroundings in which data is generated or exists. This approach focuses on understanding how data is influenced by its environment and how that context impacts the interpretation and relevance of the data. Finally, DCH will provide the means to simplify the contextual data analysis by aggregating data from different sources.

Reasoning is the process of deriving logical conclusions or understanding relationships between pieces of information based on the meanings and context associated with the data. The knowledge graph that is created by DCH can be used as a basis for analysis and reasoning tasks. Different methods can be used to perform reasoning, such as semantic, rule-based, and ML-based reasoning. Semantic reasoning involves the application of semantic knowledge and rules to make inferences, draw conclusions, or understand connections that go beyond explicit data relationships, while ML-based reasoning relies on ML approaches for the reasoning task.

The key advantages of DCH are the following:

- **Enhanced Data Contextualisation:** Knowledge graphs organise information in an interconnected way, enabling AI models to understand the relationships between different pieces of data. This

leads to a deeper understanding of the context, improving the AI's ability to generate more accurate and relevant responses.

- **Improved Information Retrieval:** By structuring data as a graph, it becomes easier for AI models to navigate and retrieve specific pieces of information quickly. This efficiency in information retrieval means that users can get answers to their queries faster, enhancing productivity.
- **Personalised User Experiences:** Knowledge graphs can help in tailoring information to the specific needs or profiles of users by understanding the connections between various data points. AI models can leverage this to provide personalised responses, improving user satisfaction and engagement.
- **Scalability and Flexibility:** As enterprises grow, so does their data. Knowledge graphs are scalable and flexible, allowing for the easy addition of new information without disrupting existing structures. This scalability ensures that AI models can continuously provide up-to-date and comprehensive answers.
- **Enhanced Decision Making:** By providing AI models with a structured understanding of enterprise knowledge, employees at all levels can make more informed decisions. Access to a broader and more accurate knowledge base can lead to better outcomes in strategy, operations, and customer service.
- **Cross-Domain Knowledge Integration:** Knowledge graphs excel in integrating information from various domains within an enterprise, breaking down silos.
- **Accessibility:** Main knowledge accessibility empowers AI models to provide insights that consider multiple aspects of the business, leading to holistic solutions.

6. Allocation of Resources

The cost of setting up and maintaining the GIANTS SharePoint is paid by the project coordinator VIF. The repository will not be updated after the end of the project. However, the files stored there will be accessible for 4 years after the end of the project. Publication costs will be covered by the respective partners/authors. Resources and efforts have been planned in the partners' budgets during the proposal phase. If the costs are eligible and planned, Open Access publication costs will be covered. Further details can be found in the Grant Agreement [2].

In its role as coordinator, VIF is responsible for setting up and updating the SharePoint, maintenance (structure, uploads, downloads, etc.), security assessment and mitigation measures, as well as access to the SharePoint and management of user requests. The quality of the data is the responsibility of the data provider/owner (in this case, the GIANTS partner). Data is handled in accordance with the FAIR principle and is made available as openly as possible by all GIANTS partners without jeopardising the intellectual property rights (IPR) of the partner(s).

The costs of the data handling tools ICOS and Data Context Hub are covered by ViF. The necessary licenses to access the data is provided by ViF as test licenses until the end of the GIANTS project.

7. Data Security, Ethics and Other Issues

7.1 Data Security

The GIANTS partners make every effort to protect the data from unauthorised use and to provide secure access to data. All shared and processed data for GIANTS is stored in a secure environment (the GIANTS SharePoint) with access rights restricted to the relevant project partners. Data transfers between partners follow the respective Articles in the CA [3] and GA [2].

7.2 Ethics

GIANTS will be conducted with the highest ethical standards and in accordance with applicable EU, international and national legislation. Ethical and legal issues are addressed in the CA [3] and GA [4] of the GIANTS project. The ethical review carried out did not identify any ethical issues. So far, no ethical issues have been identified. Further ethical issues will be added in the updated Data Management Plan as needed.

7.3 Other Issues

No other issues have been identified at this time of the project. Other issues will be added here as needed.

8. Conclusion

This Data Management Plan provides the first guidance for the data handling in the GIANTS project. The partners are aware of the regulations and recommendations that are provided by the European Commission. Workshops about proper data management will be held during the runtime of the project when it becomes necessary.

The GIANTS partners are aware of the FAIR Data principles and will act accordingly. The IPR are regulated carefully in the GA [2] and CA [3] which are signed by all partners. It is clear, that this deliverable (D1.2) provides a first version of the Data Management Plan. The updated version of the Data Management Plan will be uploaded in month 16 of the GIANTS project runtime.

9. Abbreviations

Term	Definition
API	Application Programming Interface
D&C	Dissemination and Communication
DCH	Data Context Hub
GIANTS	Green Intelligent Affordable New Transport Solutions
IPR	Intellectual property rights
PU	Public
R	Document, Report
WP	Work Package

10. References

[1] “Regulation (EU) 2016/679 regarding the protection of personal data,” [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504. [Accessed 25 June 2024].

[2] *Grant Agreement - 101138220 - GIANTS.*

[3] *Consortium Agreement - GIANTS.*

[4] “Data Contect Hub,” [Online]. Available: https://www.datacontexthub.com/. [Accessed 25 June 2024].

[5] “Cordis,” [Online]. Available: https://cordis.europa.eu/. [Accessed 25 June 2024].

[6] “GIANTS,” [Online]. Available: https://giants-project.eu/. [Accessed 25 June 2024].

[7] “Zenodo,” [Online]. Available: https://zenodo.org/. [Accessed 25 June 2024].

[8] “Kubernetes,” [Online]. Available: https://kubernetes.io/. [Accessed 25 June 2024].

[9] “MS Azure,” [Online]. Available: https://azure.microsoft.com/en-us/. [Accessed 25 June 2024].

[10] *GIANTS Deliverable D1.1 “Project Management Handbook”.*