



Deliverable D8.6 – Data Management Plan

GRANT AGREEMENT NUMBER: 101070141



GLACIATION

Project acronym: GLACIATION

Project full title: Green responsibLe privACy preservIng dAta operations

Call identifier: HORIZON-CL4-2021-DATA-01-01

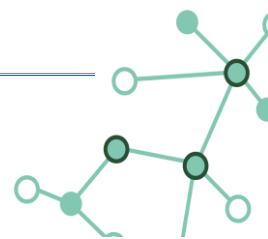
Type of action: RIA

Start date: 01/10/2022

End date: 30/09/2025

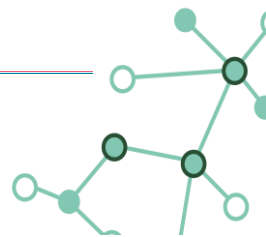
Grant agreement no: 101070141

D8.6 – Data Management Plan





Executive Summary:	D8.6 outlines the procedures and strategies that will be used to manage research data throughout the project		
WP:			8
Author(s):			Alessandro Paciaroni
Editor:			LC
Leading Partner:			The Lisbon Council
Participating Partners:			MEF, DELL, UNIMI, ENG
Version:	0.9	Status:	Final
Deliverable Type:	(R) Report	Dissemination Level:	(PU) Public
Official Submission Date:	31/03/2023	Actual Submission Date:	31/03/2023



Disclaimer

This document contains material, which is the copyright of certain GLACIATION contractors, and may not be reproduced or copied without permission. All GLACIATION consortium partners have agreed to the full publication of this document if not declared “Confidential”. The commercial use of any information contained in this document may require a license from the proprietor of that information. The reproduction of this document or of parts of it requires an agreement with the proprietor of that information.

The GLACIATION consortium consists of the following partners:

No.	Partner Organisation Name	Partner Organisation Short Name	Country
1	MINISTERO DELL'ECONOMIA E DELLE FINANZE	MEF	IT
2	EMC INFORMATION SYSTEMS INTERNATIONAL UNLIMITED COMPANY	EISI (DELL)	IE
3	HIRO MICRODATACENTERS B.V.	HIRO	NL
4	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER	LUH	DE
5	THE LISBON COUNCIL FOR ECONOMIC COMPETITIVENESS ASBL	LC	BE
6	UNIVERSITA DEGLI STUDI DI MILANO	UNIMI	IT
7	UNIVERSITA DEGLI STUDI DI BERGAMO	UNIBG	IT
8	GEIE ERCIM	ERCIM	FR
9	EURECOM	EURECOM	FR
10	SAP SE	SAP SE	DE
11	UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK	UCC	IE
12	SOGEI - SOCIETÀ GENERALE D'INFORMATICA S.P.A.	SOGEI	IT
13	LAKESIDE LABS GMBH	LAKE	AT
14	ENGINEERING - INGEGNERIA INFORMATICA SPA	ENG	IT
15	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETH	CH



Document Revision History

Version	Description	Contributions
0.1	Table of Content	LC
0.2	Version for internal review	LC
0.3	Version reviewed by the Project Coordinator	MEF
0.4	Version reviewed by Technical Coordinator	DELL
0.5	Version reviewed by third reviewers	ENG, UNIMI
0.6	Version with reviews implemented	LC
0.7	Version with final review of Technical Coordinator and Project Coordinator	DELL, MEF
0.8	Version with changes implemented by WP Leader	LC
0.9	Last version reviewed and submitted	MEF

Author

Author	Partner
Alessandro Paciaroni	LC

Reviewers

Name	Organisation
Alessio Chellini	MEF
Laura Coticoni	MEF
Francesco Pane	MEF
Delia Milazzo	ENG
Aidan O'Mahony	DELL
Pournima Sonawane	DELL
Pierangela Samarati	UNIMI

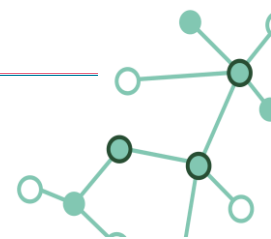
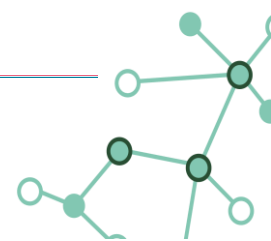




Table of Contents

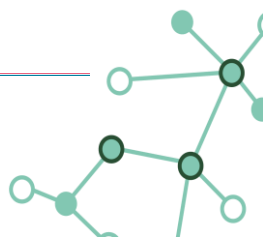
1	Introduction	9
2	GLACIATION Data Management Plan	10
2.1	Data Summary	10
2.2	Data storage, access, and security.....	13
2.2.1	Data quality	14
2.2.2	Data availability and sharing between GLACIATION partners	14
2.2.3	Archiving, preservation, and deletion of data.....	14
2.3	Making data FAIR	15
2.3.1	Making data findable, including provisions for metadata.....	15
2.3.2	Making data accessible	16
2.3.3	Making data interoperable.....	16
2.3.4	Increasing data re-use.....	16
2.4	Management of other research outputs	17
2.5	Ethics	17
3	GLACIATION Research Ethics and Compliance Protocol	18
3.1	General principles of data protection under GDPR.....	18
3.2	Data protection policy.....	18
3.3	Data protection impact assessment.....	19
3.4	Ethical issues in GLACIATION	19
4	Responsible Research and Innovation practices	20
4.1	Open science	21
4.2	Public engagement	22
4.3	Ethics and integrity.....	23
4.4	Gender equality and inclusiveness.....	23
5	Conclusions.....	25
	Annex A – DMP Questionnaire.....	27





List of Tables

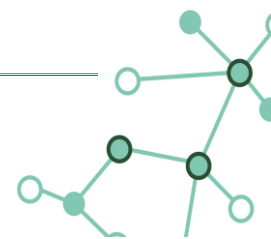
Table 1 - Categories of data used in GLACIATION 11





List of Terms and Abbreviations

Abbreviation	Description
CC0	Creative Commons 0
DKG	Distributed Knowledge Graph
DMP	Data Management Plan
DoA	Description of the Action
DPIA	Data Protection Impact Assessment
DRI	Decentralised Resource Identifier
EAB	External Advisory Board
FAIR	Findable-Accessible-Interoperable-Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
GEP	Gender Equality Plan
GLACIATION	Green responsible privacy preserving data operations
H2020	Horizon 2020
HE	Horizon Europe
ML	Machine Learning
Mxy	Month xy of the project's duration
PA	Public Administration
R&I	Research and Innovation
RRI	Responsible Research and Innovation
UCx	Use Case x
WP	Work Package





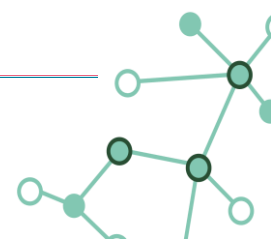
Executive Summary

This Data Management Plan (DMP) outlines the strategies and procedures that will be used to manage research data throughout the Horizon Europe (HE) project. The importance of data management in ensuring the integrity, accessibility, and long-term preservation of research data, as well as the value of incorporating Responsible Research and Innovation (RRI) principles into practices, is recognised.

The DMP describes the types of data that will be generated by the project, the methods that will be used to collect, store, and analyse this data, and the policies that will be put in place to safeguard data privacy and security. The mechanisms that will be used to ensure the proper documentation, sharing, and preservation of research data, including the use of appropriate metadata standards and data repositories, are also outlined. For the purpose of this document, a useful distinction is made between personal and non-personal data. In this regard, it is noteworthy that the project consortium is striving to use synthetic data when personal data is needed and when suitable for the objectives of the research. However, this is certainly excluded in the case of stakeholders' management.

The project team is committed to promoting open science and engaging with the public throughout the research process, as well as adhering to ethical principles and ensuring gender equality and inclusiveness. In this sense, the use of synthetic data is to be interpreted as an ethical approach pursued to avoid unnecessary use of real individuals' data while still conducting relevant research that would require personal data. The societal impact of the research will be assessed, and efforts will be made to engage stakeholders in the development of research activities.

The DMP is a lively document and will be updated and revised as necessary throughout the course of the project, and at least in an official update planned in December 2023 (M15), with the goal of ensuring that data management practices remain current and effective. By adhering to the principles of Responsible Research and Innovation (RRI), the aim is to conduct research that is not only scientifically sound but also socially responsible and impactful.





1 Introduction

The purpose of this deliverable is threefold. Firstly, it aims to provide an overview of the general categories of data that the project is expected to collect, process, and/or generate, as well as the way in which the data will be handled by partners during and after the project. This includes details on the processes used to gather the data, secure it, and make it available, as well as adherence to the FAIR Guiding Principles, which ensure that data and research outputs are findable, accessible, interoperable, and re-usable.

Secondly, the deliverable establishes the procedures that will be followed by GLACIATION partners to ensure that ethical and data protection standards are met, and appropriate measures are taken throughout the research process. This ensures that all data is handled in a secure and responsible manner.

Thirdly, the deliverable identifies preliminary RRI practices that will be relevant to the consortium's joint and individual research efforts during the project. By incorporating RRI practices into the research process, the aim is to conduct research that is not only scientifically sound but also socially responsible and impactful.

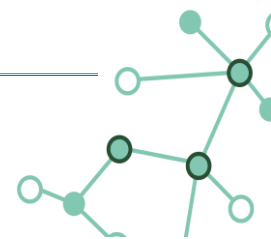
Overall, the deliverable seeks to ensure that data management practices are transparent, effective, and meet ethical standards, while also promoting responsible research and innovation. As the project progresses, the deliverable will be updated and revised to ensure that data management practices and RRI principles remain up-to-date and relevant.

Information on the data collected and processed during the GLACIATION project included in this document were gathered using the Data Management Plan (DMP) template questionnaire made available by the European Union Commission.¹ It also includes some additions according to the Consortium of European Social Sciences Data Archives (CESSDA)² and further considerations on the basis of the Practical Guide to the International Alignment of Research Data Management. The questionnaire was distributed to the consortium through the shared repository.

The document is structured as follows: the first section after the introduction, Chapter 2, is dedicated to the specific data management plan, providing information regarding data categories, provenance, storage, security, and “FAIRification”. Chapter 3 addresses more specifically research ethics and compliance. Chapter 4 consists of the considerations on RRI. Finally, conclusions are presented in Chapter 5. Annex A provides the questionnaire used as part of the work to develop this deliverable.

¹ European Commission (2021), Horizon Europe Data Management Plan Template (Version 1.0), <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>, retrieved on 03-01-2023.

² Science Europe (2021), Practical Guide to the International Alignment of Research Data Management, DOI: 10.5281/ZENODO.4915861.





2 GLACIATION Data Management Plan

A Data Management Plan (DMP) is a document that outlines how data will be collected, stored, managed, preserved, shared, and/or reused throughout a research project. It is a roadmap that helps researchers navigate the complex landscape of research data management and ensures that research data are well-documented, well-organised, and easily accessible to researchers, collaborators, and other stakeholders both during and after the research project.

Creating a data management plan is important for several reasons. First, it helps researchers to think proactively about how they will manage their data throughout the project lifecycle, which can help to reduce the risk of data loss, increase the efficiency of their research workflows, and facilitate the replication and reuse of their findings. Finally, a DMP can help to ensure that the research data are accessible and useful to others, which is essential for advancing scientific knowledge and promoting open science.

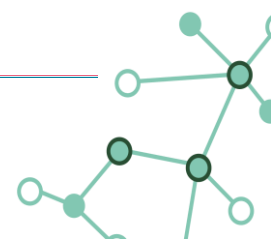
The present data management plan includes information on the following topics:

- The types of data that will be collected.
- The data collection and management, including but not limited to use cases.
- The policies and standards that will be followed for data formatting, documentation, and sharing, or fitting considerations regarding FAIR data.
- The roles and responsibilities of different stakeholders.
- The data security and backup procedures.
- The plans for data archiving and long-term preservation.

This section presents the results of the input collected from partners regarding the data management aspects of GLACIATION, it includes making data FAIR and the management of other research outputs.

2.1 Data Summary

Table 1 summarises the categories of data used in GLACIATION identified at the time of writing. It shall be noted that this is a first attempt at categorising data and their datasets, which shall by no means be considered definitive. Stakeholders' data shall be subject to Data Protection Impact Assessment (DPIA), together with pilots' data in case these will not use synthetic data.



ID	Category of data	Management	Pilots	Technology development	Dissemination
1	Stakeholders' data				X
2	Sensor data		X		
3	Log data		X		
4	Production data		X		
5	Environmental data		X		
6	Maintenance data		X		
7	Workload data		X		
8	Timestamp data		X		
9	Employees data		X		
10	Energy policy data			X	
11	Privacy policy data		X	X	
12	Usage data			X	
13	Analytics data			X	
14	Architecture data			X	
15	Management data	X			

Table 1 - Categories of data used in GLACIATION

Types, format, and size of data collected, generated, and used for the purposes of the GLACIATION project, consist of a variety of categories including but not limited to strings, floating, integer and boolean, in addition to abstract data types such as timestamp and geospatial data. Data will be stored in several formats, both structured and unstructured, including but not limited to XML, CSV, JSON, image (JPG, PNG, SVG, EPS), text (DOCX, PDF), real time streaming.

Use and reuse of data fulfils several purposes and relies on different rationales in GLACIATION, depending on the specificities of the tasks. The GLACIATION project will use historical and external data sources to optimise green energy usage within a manufacturing facility. The collected data will be stored on the GLACIATION platform, which will provide the necessary infrastructure for processing and analysing the data. The purpose of generating and reusing the data is to support the development and validation of the GLACIATION platform to reduce energy consumption, improve predictive analytics, and optimize data movement using distributed graph and AI/ML technologies. The data sources will include both renewable and non-renewable energy sources, such as solar, wind, and fossil fuels, and will be sourced continuously in real-time. Synthetic or anonymised data will be used throughout the use case to reflect the real data produced by pilots, with the management of presence/absence of its employees, but without using real personal data. The purpose of generating or reusing the data is to support the development and validation of the GLACIATION platform for the pilot, which is focused on edge-decentralised data. The data will be generated, stored, and processed by the platform owned by the partner. Publicly available data will be used as well in a specific pilot (i.e., UC3 “Privacy-preserving cross-company analytics cross-organisational”).³

³ See for example: <https://archive-beta.ics.uci.edu/donation> uses Creative Commons Attribution 4.0 International license (CC BY 4.0)

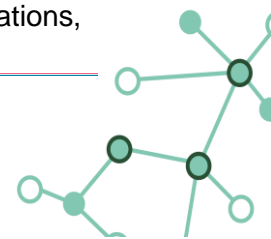


Data is also collected regarding management and coordination of distributed knowledge graphs (DKGs) to ensure energy-efficient data movement while preserving privacy, specifically in tasks 3.2, 3.3, and 7.3 (respectively “AI/ML-based workload placement”, “Swarm intelligence-based orchestration”, Integrate, validate and demonstrate use cases in integrated environments”). The origin of the data includes workload description, metadata, and knowledge within the DKGs, with collection procedures still to be defined in WP7 (“Validation and Use Cases). The first data is expected to be used in the context of WP3 (“AI-enabled Data movement Engine”), during the second year of the project.

Data is used to train an ML model for optimising job placement on a designated cluster to minimise energy consumption. The data will be collected through a telemetry mechanism installed on the cluster and owned by the responsible partner, but it will also be accessible to all other partners and the public as supplementary scientific and technical papers. The purpose of generating or reusing data is to optimise the cluster for minimising energy consumption through the training of an ML model. The data collected will include telemetry data related to the cluster's energy consumption, job placement, and other relevant metrics.

The partner who created the dataset "owns" the dataset and has the right to control access to it, but no partner will intentionally prevent another partner from accessing the dataset if it's necessary to achieve the project's objectives. Partners will have the necessary access to the data for the duration of the project. Publicly available data will be used as well.

Personal data will be collected and used for communication and dissemination activities related to the GLACIATION project. The data collected includes first and last name, email address, organisation, and role in the organisation. These unique identifiers are necessary for the project to ensure transparency and effectively engage stakeholders through personalised communication and collaboration proposals. The data will be stored securely in a relational database using high-security standards and regular backups. Participants will be asked to provide informed consent and have the option to opt-out at any time. The GLACIATION's website will use the software Matomo Analytics to track and manage usage statistics, complying with current European regulations on privacy protection and data management (GDPR). Only authorised employees will have access to the database, ensuring second-level access control. The collected data will be accessible to all GLACIATION partners, as well as the public, as a supplementary scientific/technical paper. Overall, the data will be limited to the purpose of registration and event attendance and will be deleted upon opt-out request. Personal data of employees is also collected for use cases. The data collected will not include any special categories of personal data (e.g., data revealing ethnic origins or political opinions), and will be limited to the civil servants' name and surname, gender, working area, working hours, and presence/absence from the workplace. To ensure privacy and data protection, the objective is that only synthetic or anonymised data will be used throughout the pilots. This is to prevent any potential harm or misuse of personal data, and to comply with relevant data protection regulations. The purpose of collecting this data is to enable pilot partners to develop more efficient and effective systems, with the ultimate goal of improving service delivery. Synthetic data will be used for the purposes of the use cases; the data that will be collected includes employee names and IDs, contact information such as email address, phone number, and address, job titles and functions, training and certification records, and shift schedules and working hours. The use of synthetic data ensures that the privacy of actual employees is protected, while still allowing for the analysis of workplace practices and processes. The data will be collected in compliance with legal obligations, such as health and safety regulations,





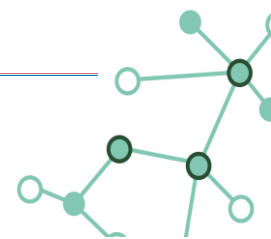
and in accordance with the legitimate interests of the employer in improving productivity and efficiency as expressed in the interested company's data policy. The data will be securely stored and accessed only by authorised personnel. Any future use of the data will require adherence to the same data protection and privacy standards. As per GLACIATION Grant Agreement: "All processing of personal data will be conducted in accordance with the provisions of: the GDPR (Regulation (EU) 2016/679) (European Union, 2016), the Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data, and the national laws applying its provisions, including those governing the acquisition of valid legal ground for processing. Data managed during the project will be processed only under the following preconditions which need to be met: (1) when the data subject has given her/his legal ground for processing; (2) when the processing is necessary for the performance of or the entering into a contract; (3) when processing is necessary for compliance with a legal obligation; (4) when processing is necessary in order to protect the vital interests of the data subject."⁴

Data collected in the context of GLACIATION may be useful to several other actors. Some may be other branches and departments of the consortium members. For instance, results achieved with the use case UC1: "Edge-decentralised data management" could be useful for use case owners and for the Italian PAs, in terms of improving performance goal in digitalisation and for implementing efficient and optimised public information system through potentiality of reduction of energy consumption and waste of processing time; this while guaranteeing privacy preservation and data security across the entire data process lifecycle. Data managed in the context of GLACIATION may be useful to other data centre management teams and Edge-Cloud infrastructure providers/operators and researchers working on similar topics. The data generated through this project with regards to the UC2 "Data-driven energy-efficient manufacturing" has the potential to be useful for several different actors within the manufacturing industry and beyond. Researchers and academics studying manufacturing processes and industrial efficiency could benefit from the data to gain insights and develop new techniques and technologies. Equipment manufacturers and vendors may also find the data useful to improve their products and services, as they can use the data to identify areas of weakness and develop solutions to improve the performance of their equipment. Other manufacturing companies may also find the data helpful in order to improve their own processes and efficiency, as they can learn from the insights that can be gained from the data. Finally, regulatory bodies or government organisations that monitor and regulate the manufacturing industry may also find the data useful for their own analyses and assessments, as it can provide them with valuable information about the state of the industry and potential areas for improvement.

2.2 Data storage, access, and security

The data created during the GLACIATION project will be stored by the GLACIATION consortium partner(s) who owns or provides the dataset. The partner(s) owning the dataset will generally control access to the dataset, and will be responsible for collecting, storing, and deleting the data. Specific circumstances arising from the use of certain technologies/solutions

⁴ Grant Agreement, part B, p. 18





and/or the nature/needs of specific activities in the context of GLACIATION may result in a different arrangement. Should circumstances that demand a different arrangement arise, appropriate solutions will be devised and duly reported. Additionally, in the circumstances where publicly available data will be used their storage, access, and security are managed by the provider. The security of the data will be ensured by implementing appropriate technical and organisational measures to prevent unauthorised access, accidental loss, or destruction of data. Access controls will be put in place to ensure that only authorised personnel can access the data, and regular backups will be taken to ensure that data is not lost due to hardware or software failure. Data encryption will also be implemented to ensure that data is protected during transmission and storage.

2.2.1 Data quality

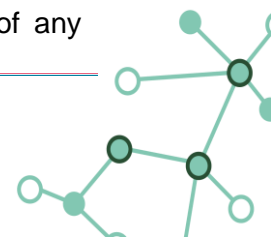
Data quality is ensured across datasets by GLACIATION consortium partners (dataset owners) by using standard practices of version control. This applies to the partners organisations' repositories and the project's repository (i.e., GitHub) when suitable. All organisations enforce standard access control and authentication mechanisms, albeit they implement them using different technologies. For instance, some partners use a combination of Drupal and MySQL and will implement a two-layers access control whereas others will use the Active Directory of their own organisation. The most intensive and extensive use of personal data is done for stakeholders' management, in this case a list of authorised individuals, akin to data stewards, will be submitted to the consortium.

2.2.2 Data availability and sharing between GLACIATION partners

Overall, personal data is not expected to be shared. Shall a need for this arise, appropriate measures will follow the execution of the activity, hence consent from the data owner will be required and information on the necessity and purpose of the data sharing will be provided. However, it shall be restated again that this is not foreseen. Data regarding the architecture are expected to be shared and accessible by partners through the project's repository, likewise GLACIATION's management data and material. Data availability and sharing between partners will be further defined in the continuation of the activities regarding the definition of concrete work arrangements during the project and will be specified in the update of this deliverable.

2.2.3 Archiving, preservation, and deletion of data

The partner(s) owning the dataset will generally control access to the dataset, and will be responsible for collecting, storing, and deleting the data. Various types of storage will be used during the project, including local/on premise/on device storage, (proprietary) distributed storage (using cloud storage as a data backend), and cloud storage (which may either be arranged and used individually by a partner (e.g., Azure, AWS, Hetzner, OneDrive, etc.). All partners implement backups according to the procedures of their organisation. Overall, data will be deleted according to industry-standard secure data deletion techniques, which will ensure that the data cannot be recovered after deletion. Backups of the data will be held for as long as is required to ensure the integrity and availability of the data in case of any





unforeseen events, such as system failure or data loss. Backups will be subject to access control and encryption when suitable and/or required by applicable legislation. The backups will ultimately be destroyed using industry-standard secure data deletion techniques. In the case of personal data, additional limitations apply, and data will be retained for up to 90 days following the end of the project. Subsequently, data will be deleted according to standards procedures.

2.3 Making data FAIR

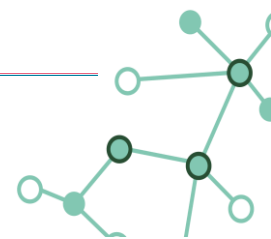
The FAIR Guiding Principles are high-level principles developed by a range of stakeholders from academia, industry, funding agencies and scholarly publishers, with the purpose of creating guidance for researchers wishing to increase the findability and, ultimately, re-usability of their data (importantly, not only by individuals, but also by machines). Given that the project is still at an early stage, it is difficult to establish which data produced during the project will have a future utility outside the project. To the extent that such potentially re-usable data are produced, compliance with the FAIR Guiding Principles will be appropriately considered. This section gives a first overview of these principles and how the GLACIATION project will take them into account. The FAIR Principles do not themselves constitute standards or specifications – instead, they are a guide to the FAIRness of data, helping researchers evaluate whether their choices are making their digital research assets FAIR. The FAIR Guiding Principles comprise four elements, which are related, but independent and separable: findability, accessibility, interoperability, and re-usability; these elements will be analysed individually in the following subsections. These principles are meant to be followed “in any combination and incrementally”, considering the context and special circumstances of each case, and can be applied not only to data, but also to non-data assets.

2.3.1 Making data findable, including provisions for metadata

The following steps lead to data being findable: (1) “(meta)data are assigned a globally unique and persistent identifier”; (2) “data are described with rich metadata”; (3) “metadata clearly and explicitly include the identifier of the data it describes”; and (4) “(meta)data are registered or indexed in a searchable resource.” Standard identifiers will be implemented, additionally the use of Decentralized Resource Identifier (DRI) is foreseen.⁵ A metadata framework will be used to identify the data types, owners and allowable use. All personal data will be managed under the GDPR, and a specific register is taken by each partner involved in their collection. More detailed description of the metadata will become available as part of the work in various project WPs and will be duly reported in subsequent versions of the present deliverable.

It shall be noticed that interoperability and specifically data knowledge graphs and metadata are a core component of the GLACIATION project which indeed aims to make this an essential aspect of its technological solution.

⁵ Trust over IP Foundation, *Decentralized Resource Identifiers in the Research Landscape*, (2021) <https://trustoverip.org/wp-content/uploads/Decentralized-Resource-Identifiers-in-the-Research-Landscape-V1.0-2021-09-23.pdf>





2.3.2 Making data accessible

Once data has been found, the next step towards potential data re-use is to know how such data can be accessed. The FAIR Guiding Principles indicate that data will be “accessible” when, (1) “(meta)data are retrievable by their identifier using a standardised communications protocol”; which (1.1) “is open, free, and universally implementable”, and (1.2) “allows for an authentication and authorisation procedure, where necessary”; and (2) “metadata are accessible, even when the data are no longer available”.⁶ Generally, data will be open access, unless the consortia or one project partner claims data to be confidential. This might be if there is a possibility for patents. In this case there might be a limited access to the data or a restriction for a certain time frame. It also might be that the data is crucial to the company’s ongoing business: algorithms. In general, data will be licenced under a public domain dedication CC0, as per the Grant Agreement unless the consortia or one of the partners claims data to be confidential. The process in this case will be defined.

2.3.3 Making data interoperable

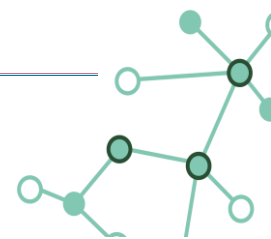
Data held by GLACIATION partners is expected to generally use widely known and community-endorsed vocabularies, standards, formats, or methodologies (e.g., json, xml, etc.), and necessary information to process the data and understand what it is by associating a publication with its metadata may be provided. At this stage of the project there are little additional information available. However, in this sense, it should be noted that a core element of GLACIATION is its work on interoperability and distributed knowledge graphs. This ensures that data interoperability is considered throughout the execution of the entire project. To be more specific, GLACIATION aims to enable interoperability across the edge-core-cloud architecture, as it is described in the GA. Therefore, making data interoperable by means of reusing existing ontologies, vocabularies and standards and thoroughly documenting new advancements on this matter are fundamental aspects of the work conducted by the consortium partners.

2.3.4 Increasing data re-use

Increasing and optimising data re-use is the ultimate goal of the FAIR principles. To ensure that data are re-usable, “(meta)data [shall be] richly described with a plurality of accurate and relevant attributes.”⁷ Documentation needed to facilitate the re-use of data will be provided by means of readme files. Additionally, in the case of datasets used as core input for other types of publications other documentation is likely to be available and made public, for instance, regarding research and analysis methodology. This goes in parallel with the more fundamental publication of data free and open access, licensed under a public license CC0. Appropriate standards will be used to facilitate re-use, such as regarding documentation on data provenance.

⁶ GO FAIR, FAIR Principles, <https://www.go-fair.org/fair-principles/>.

⁷ GO FAIR, FAIR Principles, <https://www.go-fair.org/fair-principles/>.





2.4 Management of other research outputs

Research outputs other than data are expected for GLACIATION with specific regard to code. Compliant with the provisions of RRI embedded in the Horizon Europe Programme⁸, and reflected in the project Grant Agreement, the GLACIATION consortium will operate in such a way as to ensure FAIR access to all research outputs. Exceptions to this principle shall exist only for those resources and assets that may be part of the IP of the project. In such circumstances, as it was noted in the previous sections of this chapter, specific considerations and arrangements will be undertaken. Specifically, research outputs of GLACIATION project will be published in Open Access repositories and consortium members will strive for these outputs to be pre-registered and pre-printed. Akin to datasets, research publications will be assigned persistent identifiers and characterised with metadata according to standard practices adopted by the repositories. As far as code is concerned, GitHub will be used as repository.⁹ With regards to the information regarding specific avenues for and ways of publication and dissemination of the research outputs of the GLACIATION project, at the time of writing the most complete resource is Deliverable D8.1 “Communication, Networking and Dissemination Plan and Activities”, submitted in January 2023 (M4).

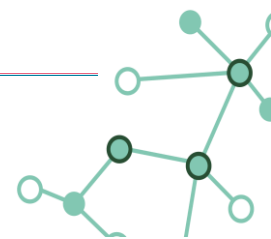
2.5 Ethics

At this stage of the project ethical considerations regarding the specific data collected, created and processed during the duration of GLACIATION are not definitive. However, GLACIATION consortium partners addressed the matter both individually and internally to their organisations and jointly with other partners. On one hand, all consortium members uptake the highest ethical standards as part of their activities and especially so within the framework of research activities like the ones conducted in GLACIATION. On the other hand, the consortium produced preliminary considerations and devised certain mechanisms to address ethical issues (such as the External Advisory Board and certain tasks in the GA as such).¹⁰ In addition to these considerations, at the moment, GLACIATION consortium partners are working on the definition of the necessary conditions and processes to use synthetic data in the project. Several partners already confirmed this decision, while others are in the process of excluding any drawbacks to this option. Nonetheless, this aspect shall not be considered as dismissed and will stay relevant throughout the project’s duration and specifically will be the focus of the next version of this deliverable as to account for any further development.

⁸ European Commission (2022), Horizon Europe (HORIZON) Programme Guide (Version 2.0), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf, retrieved on 03-01-2023.

⁹ <https://github.com/glaciation-heu>

¹⁰ GLACIATION Deliverable D1.1 “Project Management, Quality and Risk Plan” submitted in November 2022 (M2).





3 GLACIATION Research Ethics and Compliance Protocol

This section provides an overview of the research ethics in GLACIATION and the compliance with applicable regulations. This section is a combination of general principles and the input collected from partners regarding the data management aspects of GLACIATION. In future updates this section may include specific provisions as resulting from the specific tasks carried out within the project.

3.1 General principles of data protection under GDPR

According to the GDPR¹¹, the data controller, that is the natural or legal person who is in charge of determining the purposes of processing of personal data, shall ensure respect for the seven fundamental principles relating to the processing of personal data set out in Article 5 of the GDPR. Namely, the principles of (i) lawfulness, fairness and transparency; (ii) purpose limitation; (iii) data minimisation; (iv) accuracy; (v) storage limitation; (vi) integrity and confidentiality; and (vii) accountability. In other words, it shall be ensured that personal data are: (i) processed lawfully, fairly and in a transparent manner; (ii) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; (iii) adequate, relevant and limited to what is necessary in relation to the purposes for which it is processed; (iv) accurate and, where necessary, kept up to date, with every reasonable step being taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay; (v) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed; and (vi) processed in a manner that ensures appropriate security of the data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures. Finally, the data controller shall be responsible for and be able to demonstrate compliance with all the above.

3.2 Data protection policy

Pursuant to Article 24 of the GDPR, the data controller shall implement appropriate measures to ensure compliance with the regulation. Such measures are both technical and organisational. The first step to ensure compliance of a certain activity or project with the GDPR is data mapping. In other words, the identification of whether personal data is collected and processed and if so, how is it managed. Data mapping was conducted in the context of the GLACIATION project by means of the questionnaire available in Annex A. In the context of the GLACIATION project, dissemination and communication activities entail collection and processing of personal data, hence subject to GDPR compliance. Regarding this, both technical and organisational measures are taken to ensure that privacy is secured by default.

¹¹ Regulation (EU) 2016/679 (General Data Protection Regulation). Available in an easy-to-navigate format at <https://gdpr-info.eu/>





Provided that more information is available in D8.1 “Communication, Networking and Dissemination Plan and Activities”, it shall be reported here that data is collected in observance of the minimisation principle (i.e., only necessary data is collected) and data is anonymised.¹² Additionally, access control is exerted in the database to forbid access to unauthorised parties and monitor the access of the authorised ones.

3.3 Data protection impact assessment

Article 35 of the GDPR provides that a Data Protection Impact Assessment (DPIA) is mandatory in those circumstances when processing of data is likely to result in a high risk for the rights and freedoms of citizens. Specific provisions regarding when a DPIA is needed relate to those cases when sensitive data is processed on a large scale, when public areas are systematically monitored on a large scale, or when personal aspects of individuals are systematically evaluated (e.g., profiling). Admittedly, the wording of these specifications is rather generic. With this regard, the Article 29 Working Party developed a list of criteria that guide the interpretation of the regulation. Specifically, these criteria would identify circumstances in which the high risk for rights and freedoms mentioned in the GDPR is verified.¹³

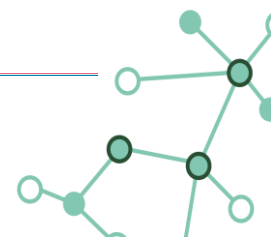
As far as the specific case of the GLACIATION project is concerned, the specific provisions of the GDPR do not indicate the need for a DPIA. In fact, use cases will demonstrate the functionalities of the technology using synthesised data. Personal data is collected only for the purposed of dissemination and communication, whose activities do not constitute a high risk to rights and freedoms of individuals. However, there are two considerations to be made. First, the list of criteria of Article 29 Working Party include the use of newer technologies. With regard to this criterion the questions arise on whether GLACIATION uses “newer” technologies. It is reasonable to argue that GLACIATION does indeed use “newer” technologies, even though it appears that the term does not elucidate on a clear set of technologies. Second, worth of mentioning in a discussion about privacy, the GLACIATION project entails a specific Work Package about data privacy, policy requirements, and protection techniques (i.e, WP4 “Protection Techniques”). Therefore, a better-informed decision on this aspect will be made in the continuation of the project and in close relation with the work conducted in WP4.

3.4 Ethical issues in GLACIATION

Throughout the project, the current existing guidelines on ethics of Horizon Europe will be rigorously applied, regardless of the country in which the research is carried out. MEF, as project coordinator will ensure that prior to any research task being undertaken the data protection and privacy regulations have been adhered to and appropriate procedures have been undertaken to ensure full legal and ethical compliance.

¹² GLACIATION Deliverable D8.1 “Communication, Networking and Dissemination Plan and Activities” submitted in January 2023 (M4).

¹³ <https://gdpr-info.eu/issues/privacy-impact-assessment/>





4 Responsible Research and Innovation practices

Promoting RRI is one of the operational objectives of Horizon Europe.¹⁴ This term broadly entails that R&I and societal values are harmonised and intertwined in a cross-fertilising relationship. On a practical level, RRI devises tools and approaches to engage the public in the research process to better align research goals and outcomes with societal needs, challenges, values, and aspirations. RRI view on harmonic R&I and society spans across several dimensions and encompasses several approaches to bring together societal values and research. Specifically, these are: open science, public engagement, ethics and integrity, gender equality and inclusiveness, and societal impact. This approach to the societal value and impact of R&I in Europe is embedded in the Horizon Europe framework through the HE Regulation¹⁵, the HE Programme Guide, and the HE Strategic Plan (2021-2024). Specifically, while the HE Regulation sets forth an impact-driven design, the HE Strategic Plan defines, precisely, the strategic orientation of the Horizon Europe Programme thereby adding a piece to the RRI puzzle of harmonic society and R&I with a strategic approach.¹⁶ In fact, Horizon Europe adopts a new monitoring approach: Key Impact Pathways.¹⁷ Following this approach, both researchers and the EU Commission can better grasp the results of Horizon funded research projects with a view on their scientific, societal, and technological impact.

According to the GA, the Horizon Europe project GLACIATION is aimed at developing a globally attractive, secure, and dynamic data-agile economy in Europe by strengthening the region's data analytics capacity through the deployment of new techniques for distributed knowledge graphs enhanced by AI technologies. The project will also focus on developing novel techniques for secure and private data sharing in common European data spaces while optimising power consumption. Through the deployment of these technologies, GLACIATION aims to develop a cloud to edge to IoT platform that can be demonstrated on several independent use cases.¹⁸ GLACIATION's intended impact is to contribute to the longer-term integration of knowledge on AI-based technologies for dynamic and predictive placement of data based on energy constraints, dynamic merging and movement of distributed knowledge graphs, novel data-centric energy monitoring frameworks, and associated metadata, and an edge-node-cloud-based platform for versatile data movement.¹⁹ These innovations will not only boost Europe's data analytics capacity but also address energy constraints by optimising power consumption, which will significantly reduce the environmental impact of data centres. Moreover, the development of novel techniques for secure and private data sharing in common

¹⁴ Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe

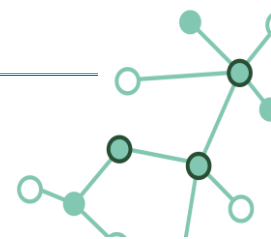
¹⁵ Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013, OJ L 170, 12.5.2021, pages 1–68.

¹⁶ Art. 50 of the HE Regulation establishes that monitoring occurs through indicators related to objectives that are, in turn, established according to impact pathways.

¹⁷ European Commission (2022), Horizon Europe programme analysis Impact assessment, evaluation and monitoring of Horizon Europe, https://research-and-innovation.ec.europa.eu/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/horizon-europe-programme-analysis_en#monitoring-horizon-europe

¹⁸ Grant Agreement, part B, p. 19

¹⁹ Grant Agreement, part B, p. 20





European data spaces will enhance the privacy and security of personal data, ensuring responsible data management in Europe. By ensuring the security and privacy of data, GLACIATION will enable individuals and organisations to benefit from the potential of data-driven innovation without fear of exploitation or misuse of personal data. In addition to contributing to the establishment of a globally attractive, secure, and dynamic data-agile economy in Europe, GLACIATION's innovations will also enable the region to be a leader in data management, setting the standards for responsible research and innovation worldwide.

The remainder of the chapter will delve deeper into the principles regarding Responsible Research and Innovation on the abovementioned dimensions and will report on how GLACIATION is committed to implement such principles as per the Grant Agreement provisions. These dimensions can be briefly characterised as follows. Open science involves making research data and findings openly available to the public to promote transparency and collaboration. Public engagement is the involvement of stakeholders and members of the public in the research process, including the design, implementation, and communication of research findings. Ethics and integrity refer to the ethical and legal standards that must be upheld in all aspects of research, including the use of human subjects, the handling of sensitive data, and the reporting of findings. Gender equality and inclusiveness involve ensuring that research activities are accessible and welcoming to all genders and social groups. Finally, assessing the societal impact of research involves understanding and addressing the potential positive and negative consequences of research on society and the environment.

4.1 Open science

Open science is premised on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the research process. It encompasses – but goes beyond – open access, which is access – free of charge for the end user – to research outputs resulting from the project. Fostering open science and ensuring visibility to the public and open access to scientific publications and research data, subject to appropriate exceptions, is a requirement and one of the operational objectives of the HE programme.²⁰ This is because open science, including open access and optimal dissemination and exploitation of knowledge, has “the potential to accelerate the advancement of knowledge by making it more reliable, efficient and accurate, easier understood by society and responsive to societal challenges [...]”²¹

Mandatory open science practices for HE-funded projects include, but are not limited to: (i) open access to peer-reviewed scientific publications relating to the results of the project, where such publications are produced and published; (ii) open access to generated research data, including those that underpin scientific publications, under the premise “as open as possible, as closed as necessary”, and taking into account the legitimate interests of the project partners; (iii) responsible management of research data in accordance with the FAIR Principles, with attention paid to long-term preservation of data; (iv) provision of information about the research outputs/tools/instruments required to validate the conclusions of scientific publications or to

²⁰ See the Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe

²¹ Preamble of the Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe





validate/re-use research data; and (v) digital or physical access to results needed to validate the conclusions of scientific publications, unless exceptions apply.

Additionally, the GLACIATION consortium will consider the following practices to the highest extent possible, as indicated in the Grant Agreement. Throughout the project, all partners will adhere to methods and procedures that ensure the early and transparent dissemination of project outcomes. Research plans will be pre-registered and registered reports will be submitted to Open Access Repositories. Additionally, methods such as pre-printing and crowdsourcing will be utilized, with consideration given to repositories and journals that accept preprints.

GLACIATION will give emphasis to guaranteeing the replicability of its outputs. In this sense, it will adopt measures that facilitate reproduction, replication, and re-use. All three processes rely on having access to the data and methods from the original study hence they rely on Open Access. With regards to this, all the research outputs that will be produced from the implementation of GLACIATION will be aligned with the Open Access and Open Science regulations of EU.

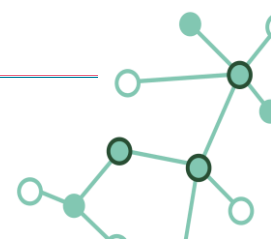
GLACIATION partners, jointly and/or individually, as appropriate, shall endeavour to take actions to respect open science requirements. By way of example, actions that have commenced or that could potentially be taken are described in the following subsections.

4.2 Public engagement

Deepening the relationship and interaction between science and society, such as through the visibility of science in society and science communication, and promoting the involvement of all societal actors, such as citizens, in co-design and co-creation processes is one of the principles and operational objectives of the HE programme. Through actions targeting public engagement, it can be ensured that society's concerns, needs, and expectations are considered, that science education is promoted, and that scientific knowledge is publicly accessible.

The GLACIATION project aims to emphasise the active involvement of the general public and non-professional scientists in all phases of the project. The involvement of end users during the implementation of pilots and validation of project outputs will be crucial to evaluate the project's results effectively. These activities will have several positive outcomes, including ensuring that the data economy and research align with the needs, expectations, and values of society, enlarging collective capabilities, research scope, and quantity and quality of data, and promoting transparency, science literacy, and public empowerment in research. The involvement of policy makers, businesses, citizens, and non-professional researchers will add value to the research outputs of the project. The project will consider best practices and knowledge from other H2020 and HE projects and initiatives to achieve the best R&I results. Overall, the dialogue model will be applied as the consortium sees fit, aiming to make it a central aspect of GLACIATION's research.²² Given the main research questions addressed in GLACIATION (i.e., related to optimising data operations across the edge to cloud continuum), the dialogue model and public engagement overall will be somewhat limited. In other words,

²² <https://rri-tools.eu/how-to-pa-public-engagement#menu-anchor-id2-content>





the specific concepts addressed in GLACIATION put a threshold to public engagement. However, openness to the public can be sought and should be sought considering the societal impact that GLACIATION's developments may have, namely through its implications for data sovereignty, the energy crisis, and perhaps also the raw materials crisis. In this sense, the consortium commits to an endeavour in engaging with stakeholders on these issues, striving to including society as it appears possible and effective. Ideally, civil society can be engaged through organisations that represent more directly citizens interests on these issues.

A multi-disciplinary approach has been opted for in the project, with partner organisations comprising a balanced variety of societal actors – universities and research centres, think tanks, business and technology associations, policy makers and businesses in various domains. All these actors will work together during the project research and implementation phases to make sure that the proposed solutions are in line with the values, needs and expectations of society.

4.3 Ethics and integrity

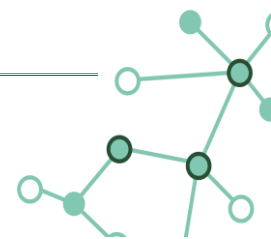
The GLACIATION consortium recognises the importance of ethics and integrity in research, and the necessity of ethical compliance, and partners are committed to carrying out their actions in accordance with the highest ethical standards and legal requirements. For this reason, an External Advisory Board (EAB) has been established to support the GLACIATION Project Management in dealing with and addressing ethics, privacy, and data protection-related issues as well as to ensure the overall quality of the innovation and project's outputs.²³ In addition, GLACIATION consortium commits to consult the EAB, shall a particular ethics concern arise, and for the evaluation of certain specific project's output. With this regard, the GA includes a project management task dedicated to “ensure ethics compliance” (i.e., T1.4).

Ethics is a central topic of research and innovation, more so regarding its conjunction with technology. With this regard GLACIATION foresees specific work on this matter. Specifically, this will occur in T3.4 (“Ethical and Trustworthy Autonomy”) and its results constitute input for D3.3 and D3.4 (respectively, “Ethical and Privacy Impact Assessment and Recommendations” intermediate and final).

4.4 Gender equality and inclusiveness

The cross-cutting principle of gender equality and inclusiveness, embedded in HE, encompasses the existence of a gender equality plan (“GEP”) within public bodies, research organisations and higher education establishments, with the aims of eliminating gender bias and inequalities and of achieving, to the extent possible, gender balance in advisory bodies and among researchers involved in projects, as well as the integration of the gender dimension into research and innovation (“R&I”).

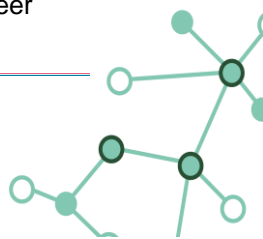
²³ GLACIATION Deliverable D1.1 “Project Management, Quality and Risk Plan” submitted in November 2022 (M2).





The project emphasises the role of the gender dimension in the economic, environmental, social, ethical, technical, and financial perspectives of the data economy, exploring whether and how specific solutions affect or concern women and men differently as end-users. GLACIATION applies a gender-sensitive approach throughout all research phases. End-users, associations, and local authorities are selected after detailed examination and pre-selection procedures to ensure a balanced group of end-users regarding gender and intersectional equality. In case of using existing data, the project also mentions potential gender biases based on the cultural or institutional contexts in which the data were generated. Finally, the project gives special attention to properly reporting gender variations in the dissemination of GLACIATION results, avoiding overemphasising gender differences, and following the SAGER publication guidelines.²⁴

²⁴ S. Heidari, T. F. Babor, P. De Castro, S. Tort & M. Curno (2016), Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use, Research Integrity and Peer Review



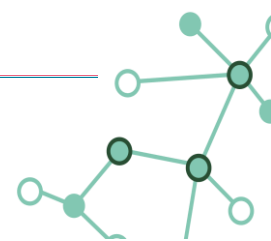


5 Conclusions

The objectives of a Data Management Plan are: first, to describe the general categories of data the project is expected to handle (collect, process and/or generate) and how. Second, to establish the procedures that shall be followed by GLACIATION partners in order to ensure that ethical and data protection standards are met, and appropriate measures are taken through the research process. Third, to preliminarily identify RRI practices that will be relevant for the research efforts during the project.

This document outlines the general categories of data expected to be handled throughout the duration of GLACIATION and provides information regarding the data type and format in which data will be stored. The main categories of data are data related to pilots, data related to project management, data related to the activities of development of the technology and data related to dissemination activities. The document proceeded with outlining data provenance, detailing different cases related to specific project activities. As far as storage, archiving, long-term preservation and the mechanisms put in place to ensure data quality and security are concerned, partners enforce the procedures established in their organisations in compliance with applicable standards. An important part of the DMP concerns consideration of the FAIR Guiding Principles – that is, how data (and other research outputs) will be made findable, accessible, interoperable and re-usable (“FAIR”). The deliverable highlights the compliance of GLACIATION consortium with the provisions outlined in the Grant Agreement and set out in the broader Horizon Framework. Importantly, based on ethical considerations, partners are working to define the use of synthetic data to ensure higher standards of privacy and security. If, however, all personal data needs to be used, all the data will be handled in compliance with the GDPR and other applicable regulations. The document concludes by outlining considerations regarding RRI principles.

The next steps with regards to data management is a more complete identification and more thorough analysis of the datasets that will be created, gathered and/or handled for GLACIATION to produce a detailed DMP for the second version of this deliverable, due in December 2023 (M15).





References and webography

Art. 50 of the HE Regulation establishes that monitoring occurs through indicators related to objectives that are, in turn, established according to impact pathways.

European Commission (2021), Horizon Europe Data Management Plan Template (Version 1.0), <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>, retrieved on 03-01-2023.

European Commission (2022), Horizon Europe (HORIZON) Programme Guide (Version 2.0), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf, retrieved on 03-01-2023.

GLACIATION Deliverable D1.1 “Project Management, Quality and Risk Plan” submitted in November 2022 (M2).

GLACIATION Deliverable D8.1 “Communication, Networking and Dissemination Plan and Activities” submitted in January 2023 (M4).

GLACIATION Grant Agreement, part B, p. 18

GO FAIR, FAIR Principles, <https://www.go-fair.org/fair-principles/>.

<https://gdpr-info.eu/issues/privacy-impact-assessment/>

<https://github.com/glaciation-heu>

<https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

European Commission (2022) Horizon Europe programme analysis. Impact assessment, evaluation and monitoring of Horizon Europe, https://research-and-innovation.ec.europa.eu/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/horizon-europe-programme-analysis_en#monitoring-horizon-europe

S. Heidari, T. F. Babor, P. De Castro, S. Tort & M. Curno (2016), Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use, Research Integrity and Peer Review

<https://rri-tools.eu/how-to-pa-public-engagement#menu-anchor-id2-content>

Preamble of the Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe

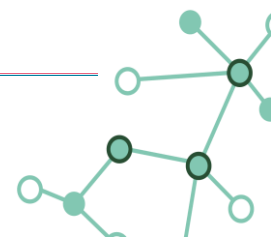
Regulation (EU) 2016/679 (General Data Protection Regulation).

Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe

Science Europe (2021), Practical Guide to the International Alignment of Research Data Management, DOI: 10.5281/ZENODO.4915861

<https://archive-beta.ics.uci.edu/donation> uses Creative Commons Attribution 4.0 International license (CC BY 4.0)

Trust over IP Foundation, Decentralized Resource Identifiers in the Research Landscape, (2021) <https://trustoverip.org/wp-content/uploads/Decentralized-Resource-Identifiers-in-the-Research-Landscape-V1.0-2021-09-23.pdf>

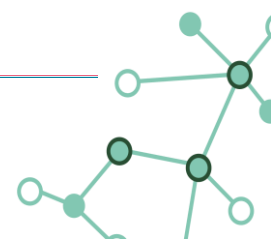




Annex A – DMP Questionnaire

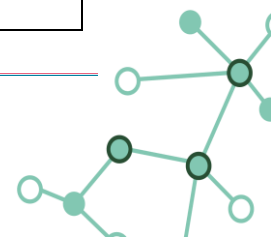
The questionnaire was built on the basis of the Horizon Europe Data Management Plan Template (Version 1.0), made available by the European Commission with the addition of “Adapt your Data Management Plan: A list of Data Management Questions based on the Expert Tour Guide on Data Management” as developed by the Consortium of European Social Sciences Data Archives (CESSDA). The questionnaire was developed for the purpose of collecting information on data management from the GLACIATION consortium. Hence, it was distributed among partners with a primary focus on the use cases partners, and the dissemination manager, then extend to all technical partners as well.

Please indicate a title for the dataset.
Please indicate the partner/s carrying out the processing operation of this dataset.
Please describe the data you will collect/generate/use. What types and formats of data will the project generate or re-use?
What is the time period covered by the dataset?
Does the dataset include personal data (i.e., data that relate to an identified or identifiable individual)? If so, please specify. Please also specify whether the dataset includes “special categories” of personal data (e.g., data revealing racial or ethnic origin, political views, religious beliefs, membership of a trade union).
Are personal data in the dataset pseudonymised? If not, please justify the answer.
Are personal data in your dataset anonymised? If not, please justify the answer.
To the extent that your dataset contains personal data, has consent been obtained for data collection or is it based on other applicable legal grounds? Have you ensured compliance with the relevant legal principles? Are data subjects in control over their data?
Will you re-use any existing data and what will you re-use it for (purpose)? In case access to data is provided by another partner or a third party, please provide information about the owner/provider of the dataset.
State the reasons if re-use of any existing data has been considered but discarded.
What is the purpose of the data generation or re-use and its relation to the objectives of the project? Please specify the task in which the processing operation will take place.
What is the origin/provenance of the data, either generated or re-used (how, where, by whom and when the data was/will be collected, produced or obtained)?
Please explain who “owns” the dataset and who will have the rights to control access to the dataset.
Identify the consortium partners (if any) having access to/using the dataset during the project. Please explain whether they will have access to the full dataset or an aggregated version.
For datasets available to other consortium partners, please provide information about conditions for





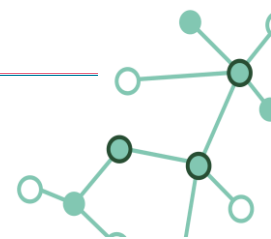
or restrictions to access to or use of the dataset (including, but not limited, to control of access).
Will personal data included in your dataset be transferred to countries outside the EU/ EEA? If so, please specify.
What is the expected size of the data that you intend to generate or re-use?
To whom might your data be useful ('data utility'), outside your project?
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
Please provide information on back-ups (e.g., how/where the data will be backed up, how often the back-up will be performed, etc.).
Please provide information about the measures adopted to ensure data integrity, data quality and confidentiality of data.
Please explain what will happen to the data after the end of the project. In doing so, please set out, among others, how/where the data will be kept (e.g., archive, trusted repository for long-term preservation and curation), for how long the data will be kept, in what format they will be kept, how they will be deleted (e.g., according to which technical standard), etc. Please also specify in your answers for how long any back-ups will be held and how they will ultimately be destroyed.
Will data be identified by a persistent identifier?
Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?
Will metadata be offered in such a way that it can be harvested and indexed?
Will the data be deposited in a trusted repository?
Have you explored appropriate arrangements with the identified repository where your data will be deposited?
Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?
Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.
If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.





Will the data be accessible through a free and standardised access protocol?
If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?
How will the identity of the person accessing the data be ascertained?
Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?
Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?
How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?
Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?
What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?
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? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?
Will your data include qualified references ²⁵ to other data (e.g. other data from your project, or datasets from previous research)?
How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?
Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?
Will the data produced in the project be useable by third parties, in particular after the end of the project?
Will the provenance of the data be thoroughly documented using the appropriate standards?
Describe all relevant data quality assurance processes.
In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g.

²⁵ A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)





<p>new materials, antibodies, reagents, samples, etc.).</p>
<p>Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.</p>
<p>Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapters in the context of the project.</p>
<p>Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?</p>
<p>What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?</p>
<p>How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)</p>
<p>Who will be responsible for data management in your project?</p>
<p>How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?</p>
<p>Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?</p>

