



# GLACIATION

Green responsible privacy  
preserving data operations

## Deliverable D8.7 – Data Management Plan

GRANT AGREEMENT NUMBER: 101070141



This project has received funding from the European Union's HE research and innovation programme under grant agreement No 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

## D8.7 – Data Management Plan

**Executive Summary:** D8.7 outlines the procedures and strategies that will be used to manage research data throughout the project

**WP:** WP8

**Author(s):** Alex Borg

**Editor:** LC

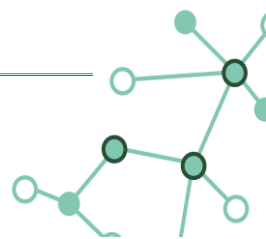
**Leading Partner:** The Lisbon Council

**Participating Partners:**

**Version:** 0.1 **Status:** Final

**Deliverable Type:** (R) Report **Dissemination Level:** (PU) Public

**Official Submission Date:** 22/12/2023 **Actual Submission Date:** 31/12/2023





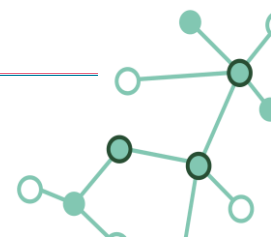
## Disclaimer

---

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	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

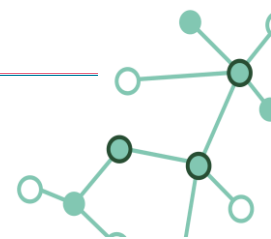
Version	Description	Contributions
0.1	Table of Content	<a href="#">LC</a>
0.2	Version for Internal Review	<a href="#">LC, MEF, ERCIM</a>
0.3	Project Coordinator Review	<a href="#">MEF</a>
0.4	Incorporation of Feedback	<a href="#">LC</a>
0.5	Final Version Ready for Submission	<a href="#">MEF</a>

### Authors

Authors	Partner
Alexander Borg	<a href="#">LC</a>

### Reviewers

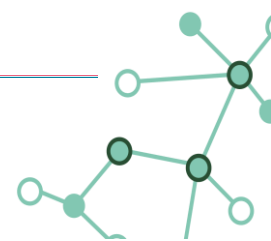
Name	Organisation
Alessandro Paciaroni	<a href="#">LC</a>
Rigo Wenning	<a href="#">ERCIM</a>
Alessio Chellini	<a href="#">MEF</a>





## Table of Contents

1	Introduction .....	9
2	GLACIATION Data Management Plan .....	11
2.1	Data Summary .....	11
2.1.1	Types of Data.....	12
2.1.2	Data Sources .....	13
2.1.3	Data Generation and Utilisation within GLACIATION: Ensuring Compliance with Privacy Standards in Research .....	15
2.2	FAIR Data .....	18
2.2.1	Make data findable, including provisions for metadata .....	18
2.2.2	Making data openly accessible.....	18
2.2.3	Making Data Interoperable .....	19
2.2.4	Increasing data re-use .....	19
3	GLACIATION Research Ethics and Compliance Protocol .....	20
3.1	Responsible Research and Innovation practices .....	20
3.2	GDPR Compliance in the GLACIATION Project.....	21
3.3	Data Protection Impact Assessment.....	22
3.4	Responsibility and resources.....	22
4	Conclusions.....	24
5	Project Datasets.....	25
5.1	WP1: GLACIATION Partners' Contact Information .....	25
5.2	WP3: Swarm Intelligence DKG Simulation .....	26
5.3	WP3: AI-Machine Learning Models .....	27
5.4	MEF Use Case: MEF Employee Dataset.....	28
5.5	DELL Use Case: Point Cloud and RGBD Robots .....	29
5.6	DELL Use Case: Robotic Control Interface .....	30
5.7	DELL Use Case: Mobile Robot-Captured Manufacturing.....	31
5.8	SAP Use Case: Data Privacy Preserving Analytics .....	32
5.9	WP8: GLACIATION Stakeholders Group .....	33
6	Data Protection Impact Assessment.....	34
6.1	Stakeholder Engagement Dataset DPIA.....	34
7	Annex GLACIATION Data Protection Impact Assessment Template.....	37
8	References and Webography .....	41

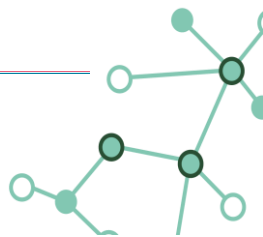




## List of Tables

---

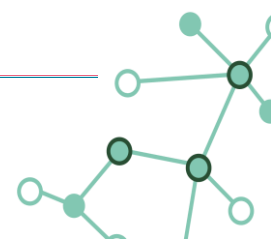
Table 1 Dataset Summary Table and Data Protection Impact Assessment Necessity ..... 12





## List of Terms and Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
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
M	Milestone
MGT	Management
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

---

As emphasised in D8.6 “Data Management Plan-Initial”, submitted in March ’23 (M6), maintaining the integrity, accessibility, and long-term preservation of research data, coupled with the integration of *Responsible, Research and Innovation* (RRI) principles, continues to be a fundamental objective of the project. This interim *Data Management Plan* (DMP) delineates the ongoing strategies and procedures aligned with *Horizon Europe* (HE) as well as other European regulations and guidelines. It offers comprehensive insights into each dataset to be utilised by the project.

The interim DMP furnishes detailed information on the data generated or used by the project, outlining methods for collection, storage, and analysis, along with policies safeguarding data privacy, security, and preservation. These measures ensure proper documentation, sharing, and preservation of research data, incorporating suitable metadata standards and data repositories. In research, synthetic data is exclusively used when personal data is necessary and aligned with research objectives, with a notable exception for stakeholder management as well as project management.

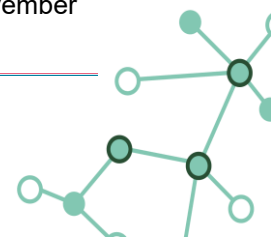
Echoing the initial DMP, the project team remains dedicated to fostering open science, public engagement, and ethical principles, prioritising gender equality and inclusivity throughout project activities. This DMP along with ethical guidelines detailed in the initial DMP (D8.6) extend beyond the minimum requirements, reflecting GLACIATION’s unwavering commitment to ethical research and personal data protection, a key objective of the project. The use of synthetic data and publicly available datasets underscores the GLACIATION project’s ethical approach to safeguarding data privacy by avoiding unnecessary use of real individuals’ data while conducting relevant research. Enhancing privacy in data movement and operations is ultimately a central goal of the GLACIATION project.

Of particular note, and in line with GLACIATION’s commitment to research ethics and data protection, an *External Advisory Board* (EAB) has been integrated into the project, comprising three experts, with one expert on “Ethics of Technologies” and another on “Privacy, Policy, and Data Management”. The Advisory Board actively participates in critical project meetings, offering consultation as needed.<sup>1</sup>

While this document represents the latest iteration of the DMP for the GLACIATION project, it remains a dynamic document subject to updates and revisions throughout the project’s duration. This ensures that data management practices stay current and effective. Emphasising the adherence to RRI principles, the project will continue to conduct research that is not only scientifically sound but also socially responsible and impactful.

---

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







# 1 Introduction

---

The GLACIATION project represents a pioneering venture at the intersection of data protection, privacy, and environmental sustainability. Far from being a zero-sum game, this project is about harmonising what are typically seen as competing imperatives. It stands as a testament to the possibility of achieving both robust data privacy and a reduced environmental footprint in the digital age.

At the heart of the GLACIATION project lies the development of an innovative metadata fabric that extends across the edge-core-cloud architecture. This initiative harnesses the power of AI to optimise data movement and operations, thereby minimising the environmental impact without compromising on privacy awareness and adherence to compliance standards. By strategically positioning analytics, GLACIATION is set to significantly lower the carbon emissions from data operations, while simultaneously bolstering data privacy.

This document presents a comprehensive account of each dataset generated and used by the GLACIATION consortium. It also delineates the data management measures implemented to not only meet but exceed the standards of Higher Education research, GDPR<sup>2</sup>, and both national and European standards, whilst protecting fundamental rights and freedoms.

The utilisation of open-source methodologies and synthetic data in the technical actions of GLACIATION negates the requirement for a *Data Protection Impact Assessment* (DPIA). This approach precludes the need for personal data analysis, systematic public space monitoring, or personal profiling.

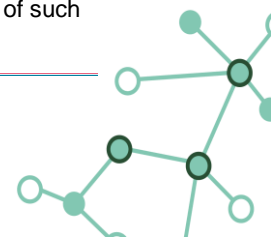
The personal data collected and used within the scope of stakeholder engagement and project management are confined to identifiers such as names and surnames. They intentionally exclude any data categories, such as genetic, ethnic, or gender information, that could potentially infringe upon individual rights and freedoms. Nonetheless, since both these activities involve personal data, a DPIA has been conducted to ensure that individual rights and freedoms will be safeguarded throughout the lifespan of the project in relation to these datasets.

The structure of this document is as follows:

- It begins with an **overview of the DMP**, summarising the types of data to be generated and utilised, the sources of this data, and the purposes for its collection, generation, and utilisation. It then provides specifics on how this data is managed in accordance with Horizon Europe guidelines, GDPR, and national and European research standards.
- Subsequently, the document explores the application of **FAIR data sharing** principles within open research environments. It outlines the essence of 'FAIRness' in research and explicates how the GLACIATION project not only complies with but is also fundamentally integrated with these principles.

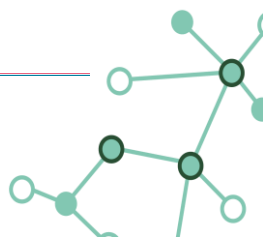
---

<sup>2</sup> REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.





- The document will then discuss **GDPR compliance** within the context of the GLACIATION project, the conditions under which a **DPIA** is applicable, and concludes with the delineation of responsibilities and resources pertaining to the Data Management Plan.
- Subsequently, the deliverable will provide a **comprehensive list of datasets**, which will include the following information: data provider and reference, purpose of data collection, relation to the project, standards and metadata, data sharing and access, data archiving and preservation, and the process for synthetic data creation.
- This will be followed by **two completed DPIAs** for the **stakeholder engagement dataset** and the **project management dataset**, illustrating how individual rights and freedoms will be safeguarded.





## 2 GLACIATION Data Management Plan

---

The *Data Management Plan* for the GLACIATION project outlines how data will be collected, stored, and managed throughout project related research. This second iteration maintains the foundational principles of the previous plan without fundamental changes. Creating a DMP is crucial for proactive data management, reducing the risk of loss, galvanising protection of personal data, enhancing research efficiency, and promoting findings' replication and reuse. The plan ensures well-documented, organised, and accessible data for researchers and stakeholders during and after the project, aligning with principles of open science.

The updated and more detailed DMP, includes:

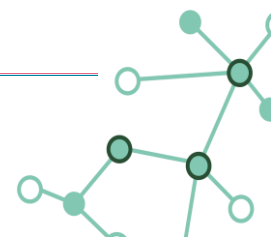
- **Data Types:** Specification of the types of data to be collected.
- **Data Collection and Management:** Description of data collection processes and management, including use cases.
- **Policies and Standards:** Adherence to policies and standards for data formatting, documentation, and sharing, with considerations for FAIR data.
- **Roles and Responsibilities:** Identification of roles and responsibilities for different stakeholders involved.
- **Data Security and Backup Procedures:** Outline of procedures ensuring data security and backup measures.
- **Plans for Data Archiving and Long-Term Preservation:** Protocols for archiving and preserving data in the long-term.
- **Input from Partners:** Results of partner input, emphasising making data FAIR and managing other research outputs in the context of GLACIATION.

### 2.1 Data Summary

---

Table 1 provides a succinct summary of the datasets employed in GLACIATION project, reflecting our continuous efforts in data management and assessment for data protection, aligning with European regulations on GDPR. It is essential to note that this process is dynamic with further datasets potentially generated and used, evolving over time in response to project needs. However, as of M15, no personal data, except for stakeholder management and project management contact details, have been generated or utilised. Specific measures have been implemented to ensure the security, privacy, GDPR compliance, and adherence to the highest ethical standards in research for these datasets, in line and going beyond HE policies.

A *Data Protection Impact Assessment* is to be conducted only if personal data is used, along with pilots' data in cases where synthetic data is not employed. Currently, personal data beyond stakeholder engagement and project has not been generated or used by the GLACIATION consortium, and there are no plans for future generation or use. However, considering that these two datasets include personal data a DPIA is necessary here.





**Table 1 Dataset Summary Table and Data Protection Impact Assessment Necessity**

Partner	Dataset	MGT	Technical	Dissemination	DPIA
<b>MEF</b>	Consortium data and project information	X			
<b>LAKE</b>	Swarm Orchestration DKG		X		
<b>UCC</b>	AI Machine Learning		X		
<b>DELL</b>	Point Cloud & RGBD Robot Navigation		X		
	Robotic Control Interface		X		
	Mobile Robot-Captured Manufacturing		X		
<b>SAP</b>	Undefined dataset for Cross Company Data Privacy Preserving Analytics		X		
<b>LC</b>	Stakeholder Engagement			X	X

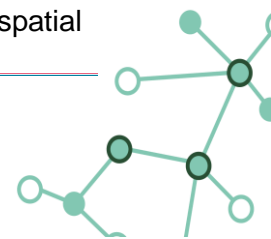
### 2.1.1 Types of Data

The data collected and processed in the GLACIATION project can be categorised into four distinct types, non-personal and personal data, as well as sensitive personal data and non-sensitive personal data:

- **Non-personal Data:** This category includes a broad spectrum of data that remains outside the GDPR purview. It features information that does not reveal an individual's identity, either directly or indirectly. Examples include aggregated datasets, synthetic datasets, anonymised records, and general environmental data.
- **Personal Data:** In line with Article 4 of the GDPR, personal data encompasses information that relates to an identifiable natural person. Identifiable individuals are those who can be recognised via specific data points, such as names, identification numbers, location data, online identifiers like IP addresses, or unique attributes pertaining to one's physical, mental, genetic, or social identity.
- **Sensitive Personal Data:** Refers to details that could reveal an individual's racial or ethnic origin, political opinions, religious beliefs, membership in a trade union, genetic and biometric data, health status, and sexual orientation.
- **Non-Sensitive Personal Data:** Encompasses any information that can identify a person, such as their name, address, date of birth, email address, and IP address, but does not include the special categories of data classified as sensitive.

The GLACIATION project is committed to handling personal data with strict adherence to current EU data protection regulations, ensuring the privacy and security of all data subjects. For each specific dataset including personal data, a DPIA will be conducted to ensure that such collection and utilisation does not threaten individual rights and freedom.

The GLACIATION project encompasses a broad spectrum of data, involving various types, formats, and sizes. It processes elementary data types like strings, floating-point numbers, integers, and booleans, as well as complex ones such as timestamps and geospatial





coordinates. The project's storage solutions are designed to support both structured data formats (such as XML, CSV, JSON, RDF) and unstructured formats, which include image files (JPG, PNG, SVG, EPS) and document files (DOCX, PDF). Additionally, GLACIATION is adept at managing real-time data streams. Upholding FAIR principles—Findability, Accessibility, Interoperability, and Reusability—in data and research is essential, ensuring that the data generated by the project can be effectively shared and reused. This data categorisation using standard formats and labels is fundamental in upholding FAIR principles, particularly interoperability.

For the activities related to communication and dissemination within the GLACIATION project, we will collect personal data as necessary. Each dataset's specific data types and handling practices will be outlined in greater detail in the subsequent sections of this document.

### 2.1.2 Data Sources

---

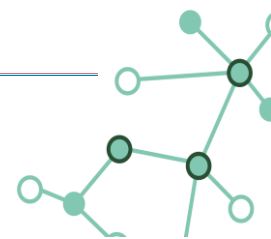
During the initial phase of the GLACIATION project, the team carefully selected a diverse array of data sources, each integral to achieving project goals. As detailed in our initial DMP (D8.6), the project taps into an extensive range of data, capturing both renewable (such as solar and wind) and non-renewable energy sources (including fossil fuels), all of which are continuously monitored in real-time. At this intermediate phase of the GLACIATION project, the team has identified and clearly defined a variety of data sources, each serving a specific purpose and playing a vital role in the project's overarching objectives:

- **WP1:**
  - **GLACIATION Consortium Partners' Contact Dataset:** Managed by MEF and stored securely on Microsoft Teams, this dataset includes contact details such as names, email addresses, and roles of consortium partners. Access is controlled and limited to verified consortium members.
  
- **WP3:**
  - **Swarm Intelligence – Billion Triples Challenge 2019 Dataset<sup>3</sup>:** Utilised for simulating the performance of our Swarm Intelligence-based Search Engine within the *Distributed Knowledge Graph* (DKG). The dataset, comprising over two billion quads from various RDF documents, is instrumental in creating a robust environment for the GLACIATION Search Engine. Access is open to all, with proper citation required.
  - **AI-Machine Learning Models – Alibaba cluster-trace-2020<sup>4</sup>:** Employed to build AI-Machine Learning models that predict future workload demands in a federated cloud system, this dataset includes GPU usage data over two months from the Alibaba PAI service. It is essential for optimising cloud system scheduling decisions. The dataset is open-source and accessible via GitHub.

---

<sup>3</sup> <https://zenodo.org/records/2634588>

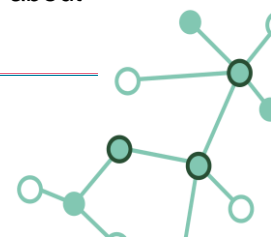
<sup>4</sup> <https://github.com/alibaba/clusterdata>





- **WP7 – MF Use Case:**
  - **MEF Employee Dataset:** Managed by Sogei & MEF and identified as DataSet\_Log\_Performance\_UC1\_TimeStamps and DataSet\_Log\_Performance\_UC1\_Reconciliation, this dataset is sourced from system and application logs. It is used for analysis, as well as training and validating models within the GLACIATION project and is accessible to all partners for machine learning training and validation of the GLACIATION platform processes. It will be preserved in the MEF repository for the next 10 years, after which it will be erased.
- **WP7 – DELL Use Case:**
  - **3D Point Cloud and RGBD of Objects in Robot Navigation in Manufacturing:** Collected to improve robotic navigation in manufacturing environments, this dataset contains detailed 3D point cloud and RGBD data. Access is regulated but available to verified researchers and industry professionals.
  - **Robotic Control Interface Configuration Dataset:** Captures a wide array of robotic control interface configurations, essential for real-time monitoring and command execution in robotic systems. The dataset is securely archived and accessible to a select professional audience.
  - **Mobile Robot-Captured Manufacturing Safety Dataset:** Aims to enhance safety monitoring in manufacturing settings using data from mobile robots. The dataset, regulated for privacy and ethical standards, contributes to safety protocol compliance and AR training.
- **WP7 – SAP Use Case:**
  - **Cross Company Data Privacy Preserving Analytics – Publicly Available Datasets:** This collection will include datasets suitable for regression tasks and will be used to test machine learning models in secure collaborative computations. It will be drawn from publicly available sources like the UCI Machine Learning Repository.
- **Work Package 8 Communication and Dissemination:**
  - **GLACIATION Stakeholder Group for Dissemination and Communication:** Comprising contact information from the GLACIATION community members, this dataset is pivotal for project communication and dissemination. Managed under GDPR standards, it ensures the ethical use of stakeholders' information and includes no data that can be considered sensitive, apart from identifiers such as names and surnames as well as contact details.

For each data source, we have defined a clear purpose, described the nature of the dataset, established the relationship to the project, and outlined standards for metadata, data sharing, access, and preservation. Detailed processes for the creation of synthetic data have been noted where applicable, ensuring all datasets align with the project's standards and methodology that are in line and go beyond HE standards. More detailed information about these datasets can be found in subsequent sections of the document.







### 2.1.3 Data Generation and Utilisation within GLACIATION: Ensuring Compliance with Privacy Standards in Research

---

Within the GLACIATION project, the generation and utilisation of data fulfil various roles, aligning with the specific requirements of distinct tasks. The project harnesses a diverse array of data sources, ranging from renewable (e.g., solar, wind) to non-renewable (e.g., fossil fuels) energy, all continuously sourced in real-time, in addition to data crucial for the development of GLACIATION technologies, project management, and stakeholder engagement.

It is pivotal to note that even in technical activities within the GLACIATION project, no dataset containing sensitive data will be used. To provide clarity, the project distinguishes between personal data, which includes non-sensitive information like names and contact details, and sensitive personal data, such as medical records or financial information. Examples of personal data in this project may include stakeholder contact information, while sensitive data, as mentioned, will not be part of the dataset utilised.

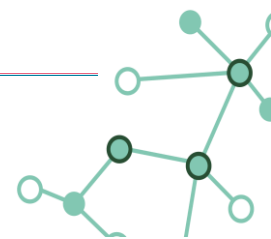
**Consortium partners' contact details** are collected by MEF, the Project Coordinator, and securely stored on MEF's Microsoft Teams server, ensuring robust data security and facilitating efficient communication and collaboration. Access to this information is strictly limited to authorised personnel, in adherence to prevailing privacy and security standards. Moreover, to reinforce the security of consortium partner data, MEF the project repository on Teams requires two-factor authentication for access. The MEF team also periodically reviews the GLACIATION mailing list and contact database to confirm the inclusion of only active project participants.

In **WP3, Task 3.2** employs the Billion Triples Challenge 2019 Dataset, referenced by Herrera, Hogan, and Käfer, for simulating the Swarm Intelligence-based Search Engine within the DKG. This extensive dataset, featuring over two billion quads from a multitude of domains, provides a comprehensive testbed for performance evaluation. Its substantial volume of partitions and files is optimally suited for integration into local Knowledge Graphs, aiding in the efficient construction of a DKG, thus creating an ideal environment for search engine testing. The dataset's open access for download, contingent on citation, aligns with the project's objectives for open data utilisation and sharing. Neither archiving nor synthetic data creation is applicable for this dataset.

In terms of **Task 3.3**, the "Alibaba cluster-trace-2020" dataset, courtesy of the Alibaba Cluster Trace Program, plays a critical role. Geared towards constructing AI and Machine Learning models, this dataset facilitates the projection of future workload demands within a federated cloud system. Covering a two-month period, it details the utilisation of over 6,500 GPUs across roughly 1,800 machines, expressly documenting AI/ML workloads managed by Alibaba's *Platform for Artificial Intelligence* (PAI) within its Machine Learning as a Service (MLaaS) offerings. This dataset is invaluable for devising models that assist cloud system orchestrators in refining scheduling decisions. When processed as a time series, it captures the aggregated workload demands comprehensively. The dataset is openly sourced and maintained in a GitHub repository<sup>5</sup>, underscoring the commitment to transparency and accessibility in support of broader research and development initiatives. The creation of synthetic data is not applicable to this dataset, emphasising its authentic real-world utility.

---

<sup>5</sup> <https://github.com/glaciation-heu>







The SOGEI and MEF dataset from **Use Case 1** is essential for the NoiPA system, meeting the edge-decentralised data management needs by monitoring performance, managing remote data collection, and aiding reconciliation. As a pivotal element of the GLACIATION project, it contributes to enhancing model training, analysis, and optimisation, whilst upholding privacy. Contrary to the initial projections in D8.6, the initial DMP, which anticipated the use of either synthetic or anonymised personal data, this Use Case will not involve the collection of personal data. Consequently, a DPIA is not required. Access to this dataset is extended to all partners, promoting collaborative progress towards privacy and energy efficiency within GLACIATION. The data will be maintained at the MEF's data centre for ten years to ensure its integrity and enable the potential for dataset reuse.

In terms of **Use Case 2**, Dell Technologies' "3D Point Cloud and RGBD of Objects in Robot Navigation in Manufacturing: Detection and Tracking" dataset is a crucial asset for enhancing robotic navigation in manufacturing contexts. It comprises high-resolution 3D point cloud and RGBD data, fundamental for precise object detection and tracking, thereby augmenting robotic interactions and efficacy in industrial settings. Linked to Use Case 2, this dataset forms the bedrock for analytics, bolstering energy-efficient manufacturing with GLACIATION's Machine Learning models. It adheres to stringent sensor and data format protocols, guarantees ethical data access via a secure portal, and is preserved in the cloud to ensure integrity and compliance. Moreover, it utilises AI simulations to generate synthetic data for diverse and lifelike scenario modelling.

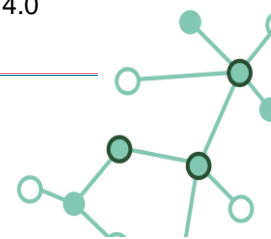
The "Robotic Control Interface Configuration Dataset" provided by DELL systematically augments the efficiency of robotic systems through real-time control and interaction analysis. This dataset is an archive of robotic interface configurations, encompassing visualisation, motion control parameters, and user interaction tools, which are critical for the refinement of robotic user interfaces. Part of actions for **Use Case 2**, it complies with robotic control norms, assuring compatibility with existing frameworks. Access is securely managed for researchers and professionals, emphasising ethical usage. Regular cloud updates keep the data integrity in line with technological advancements. The synthetic data, emulating complex control scenarios, enhances the dataset's practicality, contributing to the development of advanced, responsive robotic systems.

Dell Technologies has also curated the "Mobile Robot-Captured Manufacturing Safety Dataset" to elevate safety in manufacturing through autonomous surveillance by mobile robots employing RGB-D imaging. Vital to **Use Case 2**, the dataset includes images instrumental in honing safety protocols and augmented reality training. It conforms to robotics and imaging standards, with robust metadata facilitating detailed analytics. Securely disseminated amongst vetted professionals, the dataset is cloud-stored for enduring reliability. Synthetic data, fashioned through simulations with AI, depicts realistic manufacturing scenarios, reinforcing the dataset for pragmatic training and analytical purposes.

For **Use Case 3**, "Data Privacy Preserving Analytics" within SAP's research purview, datasets will be selected from public repositories such as the UCI Machine Learning Repository<sup>6</sup>. These datasets will serve to educate and evaluate Machine Learning models for tasks like regression, relevant to business and manufacturing insights, within secure collaborative computational settings. The data will critically appraise the protection mechanisms of GLACIATION research

---

<sup>6</sup> See for example: <https://archive-beta.ics.uci.edu/donation> uses Creative Commons Attribution 4.0 International license (CC BY 4.0)



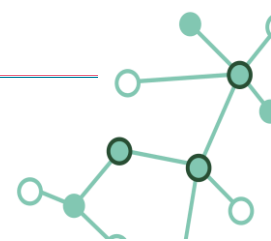


prototypes. Publicly available metadata includes subject area, task complexity, and attribute specifics, aiding in the discernment of appropriate datasets. As these datasets are publicly obtainable, their management and preservation fall under the responsibility of the providers, and SAP will comply with the extant access protocols. Presently, no synthetic data generation is envisaged; the emphasis remains on the use of extant public datasets for prototype testing.

**WP4** is intrinsically tied to the security features and privacy-preserving protocols of the GLACIATION platform, mirroring the project's profound commitment to data privacy and GDPR compliance. In the formulation and appraisal of novel protection techniques, WP4 will chiefly utilise datasets derived from the use cases within WP7. As previously mentioned, these datasets are either synthetically generated using AI technologies, thus not reliant on real personal data, or do not contain personal data. As the project approaches the finalisation of the various activities within WP4, it is acknowledged that a revisitation of the data management plan will be crucial. Adjustments will be made according to the necessities that emerge from the concluding phases of tasks vis-à-vis WP4.

**WP8's** "GLACIATION Stakeholder Group for Dissemination and Communication" dataset is critical to the project's outreach activities. Procured from the GLACIATION website, it encompasses contact details of stakeholders keen on receiving updates and engaging with the project. This dataset, containing names, email addresses, and organisational roles, is curated in accordance with GDPR guidelines. It enables direct dialogue, ensuring stakeholders remain well-informed and actively involved. Access to this dataset is exclusively reserved for the GLACIATION project team for pertinent communication purposes. In terms of data protection, the details are securely housed within Mailchimp's protected repository, consistent with GDPR compliance, and will be maintained for the project's duration or until a stakeholder chooses to opt out.

To ensure compliance and aim to surpass HE standards, GDPR, as well as national European standards and guidelines, the datasets generated and utilised, along with the methodologies adopted, are underpinned by a series of actions. These include utilising publicly available datasets that already meet HE compliance standards as well as generating synthetic data using AI to avoid any potential indirect collection of personal data. The only personal data gathered and stored are used strictly for project management and stakeholder engagement, are maintained with the highest standards of security and include only identifiers such as names and surnames, email address, deliberately omitting gender, genetic, ethnic, or other markers that could pose a risk to individual rights and freedoms.





## 2.2 FAIR Data

---

The FAIR Guiding Principles<sup>7</sup>, developed collaboratively by stakeholders from academia, industry, funding agencies, and scholarly publishers, offer high-level guidance for researchers aiming to enhance the findability and reusability of their data, catering to both individuals and machines. While it is challenging to predict the future utility of project-generated data at this early stage, the GLACIATION project remains committed to complying with FAIR Principles for potentially reusable data. This section provides an overview of these principles in line with the current phase of the project: the intermediate phase.

Not serving as standards, the FAIR Principles act as a guide for assessing the FAIRness of data, comprising four independent elements: findability, accessibility, interoperability, and reusability. This deliverable will delve into each element individually, emphasising their flexible and continued incremental application based on context and circumstances, not limited to data but extending to non-data assets as well.

### 2.2.1 Make data findable, including provisions for metadata

---

Ensuring data findability involves the following key steps: (1) assigning globally unique and persistent identifiers to (meta)data; (2) providing comprehensive descriptions with rich metadata; (3) explicitly incorporating the data identifier within the metadata; and (4) registering or indexing (meta)data in a searchable resource. The implementation of standard identifiers, including the use of *Decentralised Resource Identifier* (DRI)<sup>8</sup>, is anticipated. A metadata framework categorises data types, identify owners, and specify permissible use. GDPR guidelines govern the management of all personal data, with each partner maintaining a dedicated register for data collection. Further details on metadata, including types and ownership, has been progressively revealed through project work in various WPs and reported in subsequent deliverable versions.

Interoperability, especially through data knowledge graphs and metadata fabrics, forms a central component of the GLACIATION project. The project prioritises these aspects, aiming to integrate them seamlessly into its technological solution.

### 2.2.2 Making data openly accessible

---

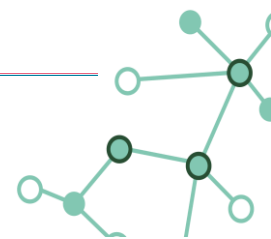
Once data is located, the subsequent step towards potential data reuse involves understanding how to access it. As per the FAIR Guiding Principles, data is deemed "accessible" when, (1) "(meta)data can be retrieved by their identifier using a standardised communications protocol"<sup>9</sup>; specifically, (1.1) "an open, free, and universally implementable" protocol and (1.2)

---

<sup>7</sup> Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

<sup>8</sup> 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>

<sup>9</sup> GO FAIR, FAIR Principles, <https://www.go-fair.org/fair-principles/>.





one that "allows for an authentication and authorisation procedure, where necessary"; and (2) "metadata remains accessible, even when the data is no longer available." Generally, data has been openly accessible unless the consortium or a project partner designates it as confidential, which may occur in cases of patent possibilities. Limited access or a temporal restriction have applied in such instances. Additionally, data crucial to a company's ongoing business, such as algorithms, may warrant restricted access. In most cases, data has been licensed under a public domain dedication CC0, in line with the Grant Agreement, unless confidentiality is asserted by the consortium or a partner. Protocols for such cases have been clearly defined.

### 2.2.3 Making Data Interoperable

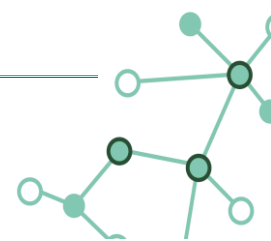
---

The Data held by GLACIATION partners utilises widely known and community-endorsed vocabularies, standards, formats, or methodologies (e.g., JSON, XML). Ultimately, a core element of the GLACIATION project is its focus on interoperability and DKGs, ensuring ongoing consideration for data interoperability throughout the project's execution. Specifically, GLACIATION aims to enable interoperability across the edge-core-cloud architecture, as outlined in the GA. Thus, achieving data interoperability through the reuse of existing ontologies, vocabularies, and standards, coupled with documenting new advancements, remains fundamental to the GLACIATION project overall and the particular work conducted by consortium partners.

### 2.2.4 Increasing data re-use

---

In alignment with the initial data management plan, this interim plan adheres to the principle of reusing data as outlined by the FAIR principles. The goal of increasing and optimising data re-use is prioritised, emphasising that "(meta)data [shall be] richly described with a plurality of accurate and relevant attributes". Documentation crucial for facilitating data re-use is provided through readme files. For datasets serving as core input for various publications, additional documentation, particularly concerning research and analysis methodology, is to be available and made public. This commitment aligns with the overarching approach of publishing data freely and openly, licensed under a public license CC0. To ensure reusability, appropriate standards, especially those related to documentation on data provenance, are diligently followed.





## 3 GLACIATION Research Ethics and Compliance Protocol

---

This section offers an overview of the research ethics in the GLACIATION project and its compliance with applicable regulations, as implemented up to M15 of the project. It synthesises general principles with contributions from partners on the data management aspects of GLACIATION. Additionally, it outlines specific actions taken to uphold the highest ethical standards and data management practices that have been diligently followed. For a more comprehensive description of the research ethics and compliance protocols employed by the GLACIATION project, please refer to the initial DMP (D8.6).

### 3.1 Responsible Research and Innovation practices

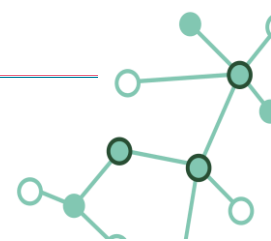
---

GLACIATION is steadfast in its commitment to ethics and legal compliance, exemplified by the establishment of an EAB dedicated to addressing ethics, privacy, and data protection concerns. The consortium actively engages the EAB for ethics-related matters and evaluations of specific project outputs, as well as in important consortium calls – as project General Assemblies and events - ensuring rigorous compliance with ethics tasks outlined in the Grant Agreement (T1.4 “Ensure ethics compliance”). Further details on ethics and responsible research practices specific to GLACIATION are comprehensively documented in D8.6, the initial DMP.

Ethical considerations are woven into the fabric of GLACIATION's research and innovation endeavours, with specific emphasis on technology-related ethics explored in T3.4 “Ethical and Trustworthy Autonomy” and detailed outcomes presented in D3.3 “Ethical and Privacy Impact Assessment & Recommendations – intermediate” and D3.4 “Ethical and Privacy Impact Assessment & Recommendations – final”. The project also aligns with gender equality principles, exemplified by the incorporation of a *Gender Equality Plan* (GEP), and is deeply committed to promoting RRI.

GLACIATION envisions and works towards creating a dynamic, secure, and globally attractive data-agile economy in Europe. Central to this vision is the development of innovative techniques for secure, private data sharing, as well as the optimisation of power consumption to mitigate environmental impacts. Detailed ethical guidelines, methodologies, and compliance measures are outlined in Deliverable D8.6, ensuring that GLACIATION remains at the forefront of responsible and ethical research practices.

Moreover, GLACIATION's commitment extends to addressing the gender dimension in the economic, environmental, social, ethical, technical, and financial realms of the data economy. Rigorous gender-sensitive approaches are applied throughout the project's research phases, ensuring balanced representation in end-user selections and acknowledging potential biases





in existing data. The project adheres to SAGER guidelines for reporting gender variations in dissemination, fostering a culture of inclusivity<sup>10</sup>.

For an in-depth understanding of GLACIATION's ethical and responsible research practices, stakeholders can refer to Deliverable D8.6, the initial DMP. This comprehensive document serves as a comprehensive guide to the ethical foundations and methodologies embedded within the project.

## 3.2 GDPR Compliance in the GLACIATION Project

---

As outlined in D8.6, the initial data management plan, has been consistently applied throughout the GLACIATION project up to the drafting of D8.7, the interim DMP, and will continue to be rigorously enforced.

Under the GDPR, the data controller—whether a natural or legal person—is tasked with determining the purposes and means of personal data processing and must adhere to the seven fundamental principles outlined in Article 5 of the GDPR. These are:

1. Lawfulness, fairness, and transparency;
2. Purpose limitation;
3. Data minimisation;
4. Accuracy;
5. Storage limitation;
6. Integrity and confidentiality;
7. Accountability.<sup>11</sup>

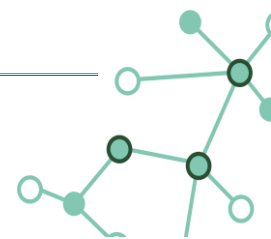
Essentially, it is crucial to ensure that personal data is processed lawfully, fairly, and transparently, collected for specified, explicit, and legitimate purposes, and is not further processed in a manner incompatible with those purposes. It should be adequate, relevant, and limited to what is necessary for processing purposes, accurate and kept up to date, with every reasonable step taken to promptly erase or rectify inaccuracies. Data should be retained only as long as necessary for the processing purposes and be processed securely to protect against unauthorised or unlawful processing, accidental loss, destruction, or damage, employing suitable technical or organisational measures. Demonstrating compliance with these principles is the responsibility of the data controller.

MEF, as the Project Coordinator, continues to ensure strict adherence to GDPR in all project activities, according to the frameworks established in D8.6 and D8.7, formulated by LC based on contributions from all partners. The GLACIATION consortium uses and stores only personal data pertaining to stakeholder management and project management. This dataset includes identifiers such as first and last names of stakeholders, email addresses collected with consent and providing an option to opt-out at any time. None of the personal data gathered, stored, and used in this regard contains any data that can be considered sensitive.

---

<sup>10</sup> 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

<sup>11</sup> Grant Agreement, part B, p. 18







Despite the project's central aim of enhancing data management and movement for a greener and more privacy-aware framework via the edge-core-cloud architecture, in line with GDPR, the technical activities of the project will utilise publicly available or synthetic datasets. This approach ensures compliance with GDPR, maintains the highest ethical standards, and obviates the need for a DPIA, as indicated earlier in the document. The methodology for generating synthetic data for each relevant dataset, which confirms no indirect collection or usage of personal data, is detailed in the dataset section below.

### 3.3 Data Protection Impact Assessment

---

Article 35 of the GDPR requires a DPIA when data processing is likely to result in a high risk to the rights and freedoms of individuals. Criteria for necessitating a DPIA include the large-scale processing of sensitive data, systematic monitoring of public areas, or systematic evaluation of personal aspects, such as profiling. The Article 29 Working Party has elaborated criteria to interpret these regulations, pinpointing situations indicative of the high risk delineated in the GDPR. As highlighted throughout this document, the GLACIATION project will not involve large-scale processing of sensitive data, systematic monitoring of public areas, nor the evaluation of personal aspects such as profiling.

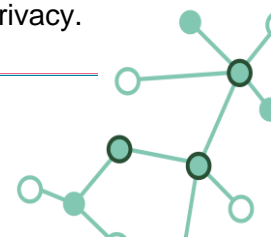
In the context of the GLACIATION project, specific provisions under the GDPR do not indicate the necessity for a DPIA. As initially laid out in D8.6, the use cases continue to demonstrate technology functionalities using synthetic data and anonymised data in Use Case 1, ensuring there is no indirect extraction of personal data from the employee dataset, thus obviating the need for a DPIA. Furthermore, within WP3, only publicly available datasets that pre-emptively comply with GDPR are utilised and which do not contain any personal data. As detailed in the dataset section below, the process for synthetic data creation also utilises AI and other techniques to eliminate any risk of direct or indirect collection of personal data.

The sole personal data to be collected and used by the project is within the realm of stakeholder engagement and project management, presenting no risks to individual rights and freedoms. The datasets include only identifiers such as names and surnames, excluding sensitive details like ethnic, gender, or genetic information, which carry a particular risk of infringing upon individual rights and freedoms. These datasets will not be used for systematic monitoring of public areas or for the systematic evaluation of personal characteristics. Moreover, this information is securely stored in the MEF Teams repository, which requires two-factor authentication for access and is restricted to consortium partners for project-related activities. These datasets will be deleted upon completion of the project.

### 3.4 Responsibility and resources

---

As the WP8 and Deliverable leader for D8.6 and D8.7, LC holds a distinctive responsibility to ensure GLACIATION's adherence to legal and ethical data protection and research ethics standards. Although the LC plays a crucial role, the primary responsibility lies with the Project Coordinator, MEF, to strictly ensure partners follow the standards set out in this document and in D8.6. At the same time, partners are responsible for the fully adherence to the GDPR principles, and to the project rules about the respect of ethics, security and privacy.

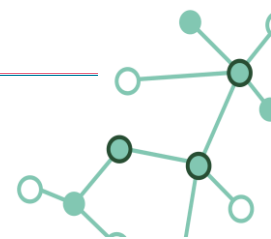






Furthermore, it is essential for every GLACIATION partner to maintain their commitment to the policies and guidelines outlined in this DMP and D8.6.

Fundamentally, the creation, management, and storage of datasets must be in strict alignment with legislation, as detailed in D8.6 and this document. MEF, as the project coordinator and as mandated in Task 1.4 “Ensure ethics compliance” of the Grant Agreement, is tasked with the effective implementation of the measures outlined in these deliverables, ensuring adherence to data and ethical guidelines. Task 1.4 requires that MEF not only meets the highest ethical standards and ensures data security and privacy but also aims to exceed these benchmarks. This commitment necessitates ongoing monitoring and assessment to guarantee that all partners embrace and consistently apply these guidelines and procedures in all project activities. Thus, adjustments to this DMP will be made as necessary, emphasising the dynamic nature of this document.





## 4 Conclusions

---

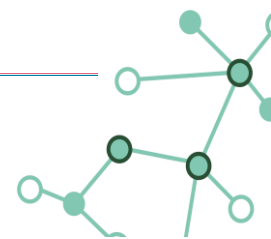
This document has meticulously detailed the datasets utilised within the GLACIATION project, demonstrating a steadfast commitment to not only meeting but surpassing the standards of delineated in the HE programme, GDPR, and both national and European research standards and guidelines. The project adheres to these rigorous protocols while maintaining an unwavering dedication to upholding individual rights and freedoms.

A DPIA has been deemed unnecessary for a dataset used in technical dimensions of the GLACIATION project. This determination rests on two pillars: the creation and utilisation of synthetic data and the use of publicly available datasets that inherently comply with privacy standards.

The sole personal data collected are confined to the realms of stakeholder engagement and project management, avoiding any large-scale processing of sensitive data, systematic monitoring of public areas, or profiling. However, considering that these datasets include personal data, a DPIA is necessary to demonstrate that individual rights and freedoms will be safeguarded. These two DPIAs can be found in the subsequent sections of this document.

As we move forward, LC will continue to provide support to MEF in the implementation of the data management measures delineated in this document, as well as those outlined in D8.6. However, it is crucial that all partners engage with the content of this document and D8.6 comprehensively to ensure these measures are fully realised.

Recognising the dynamic nature of the GLACIATION project, this document will continue to be a 'living' resource, subject to adaptation and refinement in response to the evolving needs of the project. The GLACIATION consortium is committed to ongoing vigilance in data management practices to ensure our research remains at the forefront of innovation, privacy, and sustainability.





## 5 Project Datasets

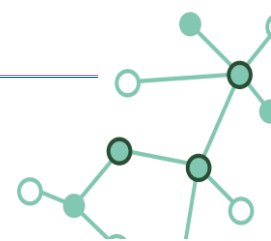
---

The upcoming section outlines datasets to be utilised in the project, providing a comprehensive list and assessing their compliance with FAIR data management principles outlined by the European Commission, all within the context of GDPR and ethical standards in research, as pertinent to Horizon Europe projects.

### 5.1 WP1: GLACIATION Partners' Contact Information

---

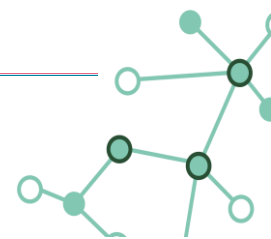
<b>Name</b>	<b>Description</b>
<b>Dataset name</b>	<i>GLACIATION Consortium Partners Contact Dataset</i>
<b>Data provider and reference</b>	<i>The dataset is maintained by MEF and stored on Microsoft Teams for enhanced security.</i>
<b>Purpose of Data Collection</b>	<i>To facilitate secure and efficient communication and collaboration among consortium partners.</i>
<b>Dataset Description</b>	<i>The dataset contains contact information of consortium partners, including names, email addresses, and roles, stored securely on MEF Microsoft Teams repository and managed by MEF, the Project Coordinator.</i>
<b>Relation to the project</b>	<i>This dataset is crucial for ensuring coordinated efforts and communication within the consortium.</i>
<b>Standards and metadata</b>	<i>All data handling complies with the applicable privacy and security standards and in turn HE standards, with access restricted to authorised personnel only.</i>
<b>Data sharing and access</b>	<i>Access to this dataset is strictly regulated, with entry requiring two-factor authentication to ensure data security.</i>
<b>Data archiving and preservation</b>	<i>The contact information is regularly verified by the MEF Team to maintain relevance and accuracy, with access facilitated through a secure and compliant repository. The dataset will be deleted upon finalisation of the GLACIATION project.</i>
<b>Process for synthetic data creation</b>	<i>Not applicable</i>





## 5.2 WP3: Swarm Intelligence DKG Simulation

Name	Description
<b>Data provider and reference</b>	<i>Billion Triples Challenge 2019 Dataset</i>
<b>Data provider and reference</b>	<i>Herrera, JM., Hogan, A., Käfer, T. (2019). BTC-2019: The 2019 Billion Triple Challenge Dataset. In: Ghidini, C., et al. The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science(), vol 11779. Springer, Cham. <a href="https://doi.org/10.1007/978-3-030-30796-7_11">https://doi.org/10.1007/978-3-030-30796-7_11</a></i>
<b>Purpose of Data Collection</b>	<i>This dataset is a large dataset we are using for the simulation of our Swarm Intelligence-based Search Engine within DKG. We measure the performance of the Search Engine using this dataset.</i>
<b>Dataset Description</b>	<i>The Billion Triple Challenge (BTC) 2019 Dataset emerges from an extensive RDF crawl carried out using LDspider between December 12, 2018, and January 11, 2019, accepting RDF/XML, Turtle, and N-Triples formats. The dataset comprises 2,155,856,033 quads, originating from 2,641,253 RDF documents spanning 394 top-level domains. In the consolidated RDF graph, formed by merging the data, there are 256,059,356 distinct triples. These quads or triples encode information with 38,156 unique predicates and instances of 120,037 unique classes. The fourth element in the quads designates the source location of the associated triple parsed from the respective web document.</i>
<b>Relation to the project</b>	<i>The dataset is useful for us, because it already has a great number of partitions, different files containing data, which can directly be loaded to a local Knowledge Graph. And the Distributed Knowledge Graph then can be built using these local Knowledge Graphs very easily, simulating a good environment for our Search Engine.</i>
<b>Standards and metadata</b>	<i>See the reference, the origin of the dataset.</i>
<b>Data sharing and access</b>	<i>The data can be downloaded and used by anyone. It only needs to be cited.</i>
<b>Data archiving and preservation</b>	<i>Not applicable.</i>
<b>Process for synthetic data creation</b>	<i>Not applicable.</i>

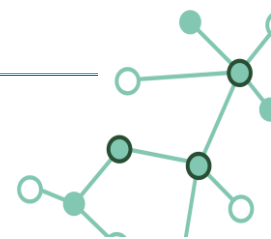




## 5.3 WP3: AI-Machine Learning Models

---

<b>Name</b>	<b>Description</b>
<b>Dataset name</b>	<i>Alibaba cluster-trace-2020</i>
<b>Data provider and reference</b>	<i>Alibaba Cluster Trace Program (<a href="https://github.com/alibaba/clusterdata">https://github.com/alibaba/clusterdata</a>)</i>
<b>Purpose of Data Collection</b>	<i>Build AI-Machine Learning models to predict the future workload demand in a federated cloud system</i>
<b>Dataset Description</b>	<i>The dataset includes over 6500 GPUs (on ~1800 machines) in a period of 2 months. It describes the AI/ML workloads in the MLaaS (Machine-Learning-as-a-Service) provided by the Alibaba PAI (Platform for Artificial Intelligence) on GPU clusters.</i>
<b>Relation to the project</b>	<i>The models build upon the dataset are useful to predict the future workload demand incoming to the cloud system and will be useful to the orchestrator to optimise scheduling decisions</i>
<b>Standards and metadata</b>	<i>The dataset is pre-processed as a time series of aggregated workload demand.</i>
<b>Data sharing and access</b>	<i>The dataset is open sourced and accessible from a GitHub repository</i>
<b>Data archiving and preservation</b>	<i>The dataset is open sourced and accessible from a GitHub repository</i>
<b>Process for synthetic data creation</b>	<i>Not applicable</i>

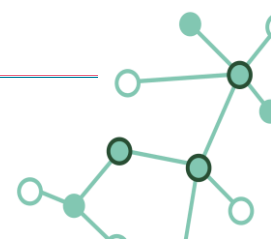




## 5.4 MEF Use Case: MEF Employee Dataset

---

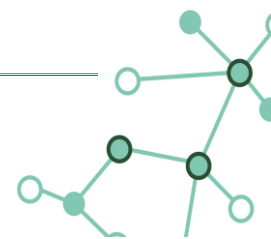
<b>Name</b>	<b>Description</b>
<b>Data provider and reference</b>	<i>Data Provider: Sogei &amp; MEF. Dataset Identifier: DataSet_Log_Performance_UC1_TimeStamps; DataSet_Log_Performance_UC1_Reconciliation</i>
<b>Data provider and reference</b>	<i>Source of the Data: System and application log.</i>
<b>Purpose of Data Collection</b>	<i>The dataset is designed to monitor system status and performance in relation to the application's workflow for remote data collection. It facilitates both daily and monthly data reconciliation. Additionally, this dataset is utilised for analysis, as well as for training and validating models.</i>
<b>Dataset Description</b>	<i>The dataset pertains to the performance metrics of the NoiPA system and the tracking of data movement.</i>
<b>Relation to the project</b>	<i>Dataset will be used for analysis purpose and to optimise workload placement and reduce energy consumption, featuring GLACIATION ML models and services.</i>
<b>Standards and metadata</b>	<i>Tracing data will be collected in Text format. Any other specification will be specified in a later state of UC1 deployment</i>
<b>Data sharing and access</b>	<i>Datasets will be generated from non-production systems, or, where this is not applicable, it is assumed that production data does not contain personal data with ethical or legal constraints on sharing. All partners will have full access to these datasets for machine learning training and for validating the GLACIATION platform process.</i>
<b>Data archiving and preservation</b>	<i>The dataset will be stored in the project repository and maintained in the MEF data centre for the next 10 years. After this period, the data will be erased.</i>
<b>Process for synthetic data creation</b>	<i>N/A, dataset will be collected using tracing services.</i>





## 5.5 DELL Use Case: Point Cloud and RGBD Robots

Name	Description
<b>Dataset name</b>	<i>3D POINT CLOUD AND RGBD OF OBJECTS IN ROBOT NAVIGATION IN MANUFACTURING: DETECTION AND TRACKING</i>
<b>Data provider and reference</b>	<i>Dell Technologies</i>
<b>Purpose of Data Collection</b>	<i>The primary purpose of this data collection is to enhance robotic navigation systems in manufacturing environments by providing detailed 3D point cloud and RGBD data for object detection and tracking. This dataset aims to improve the accuracy and efficiency of robotic movements and interactions with various objects in industrial settings.</i>
<b>Dataset Description</b>	<i>This dataset comprises high-resolution 3D point cloud and RGBD data of various objects encountered in manufacturing environments. It includes diverse object shapes, sizes, and textures, captured under different lighting conditions to mimic real-world scenarios, aiding in the development of robust navigation algorithms for manufacturing robots.</i>
<b>Relation to the project</b>	<i>This dataset is related to Use Case 2 and will be used as the data for which the analytics is run on</i>
<b>Standards and metadata</b>	<i>Standards and metadata encompass specific sensor specifications like resolution and depth accuracy, data formats for RGB-D output, calibration details including intrinsic and extrinsic parameters, and SDK-provided frame metadata such as timestamps and sensor settings.</i>
<b>Data sharing and access</b>	<i>Access to the dataset is regulated through a secure online portal, primarily available to verified researchers and industry professionals, with strict adherence to privacy and ethical use standards.</i>
<b>Data archiving and preservation</b>	<i>The dataset is securely archived in a cloud-based system with regular updates and backups, ensuring long-term data integrity and compliance with evolving technological standards.</i>
<b>Process for synthetic data creation</b>	<i>Synthetic data is generated using AI-driven simulation tools, creating realistic object models and scenarios, followed by rigorous validation to ensure accuracy and real-world applicability.</i>

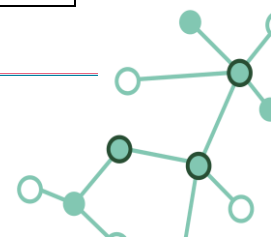






## 5.6 DELL Use Case: Robotic Control Interface

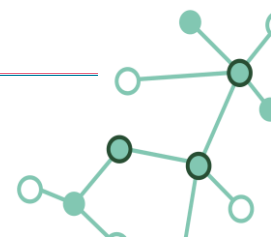
Name	Description
<b>Dataset name</b>	<i>Robotic Control Interface Configuration Dataset</i>
<b>Data provider and reference</b>	<i>Dell Technologies</i>
<b>Purpose of Data Collection</b>	<i>The purpose of this data collection is to systematically capture and analyse robotic control interface configurations, facilitating the optimisation of real-time monitoring, command execution, and interaction in robotic systems.</i>
<b>Dataset Description</b>	<i>This dataset comprises a comprehensive collection of robotic control interface configurations, including real-time visualisation settings for range and IMU data, command publishing parameters for motion and servo control, interactive indicators for user input, and monitoring tools for battery status. It encapsulates layout specifications, camera and 3D feedback configurations, and custom scripting options, serving as a valuable resource for developing and refining user interfaces in robotic system control.</i>
<b>Relation to the project</b>	<i>This dataset is related to Use Case 2 and will be used as the data for which the analytics is run on</i>
<b>Standards and metadata</b>	<i>The dataset adheres to standards and protocols relevant to robotic control systems, encompassing data formats, communication protocols, and interface design guidelines. Metadata includes detailed attributes of interface components such as plot configurations, command topics, visual feedback parameters, and interactive control elements, along with layout specifications and global variables, ensuring comprehensive documentation and compatibility with existing robotic software frameworks.</i>
<b>Data sharing and access</b>	<i>Access to the dataset is regulated through a secure online portal, primarily available to verified researchers and industry professionals, with strict adherence to privacy and ethical use standards.</i>
<b>Data archiving and preservation</b>	<i>The dataset is securely archived in a cloud-based system with regular updates and backups, ensuring long-term data integrity and compliance with evolving technological standards.</i>
<b>Process for synthetic data creation</b>	<i>Synthetic data for this dataset is generated through simulations that replicate a wide range of robotic control scenarios, using advanced algorithms to model interface interactions, sensor feedback, and command responses. This process ensures a diverse and realistic representation of various operational environments, contributing to robust and comprehensive dataset development.</i>





## 5.7 DELL Use Case: Mobile Robot-Captured Manufacturing

Name	Description
<b>Dataset name</b>	<i>Mobile Robot-Captured Manufacturing Safety Dataset</i>
<b>Data provider and reference</b>	<i>Dell Technologies</i>
<b>Purpose of Data Collection</b>	<i>To automate safety monitoring and compliance auditing in manufacturing environments using mobile robots equipped with RGB-D cameras, and to assist in augmented reality training and insurance risk assessments.</i>
<b>Dataset Description</b>	<i>This dataset comprises RGB-D images captured by mobile robots in various manufacturing settings, documenting workers' use of safety gear and compliance with safety protocols. It's tailored for applications in automated safety inspections, AR training, and regulatory compliance analysis.</i>
<b>Relation to the project</b>	<i>This dataset is related to Use Case 2 and will be used as the data for which the analytics is run on.</i>
<b>Standards and metadata</b>	<i>Adhering to robotics and imaging standards, the dataset includes metadata on robot paths, camera angles, and environmental conditions, along with detailed annotations of safety equipment usage and worker activities.</i>
<b>Data sharing and access</b>	<i>Access to the dataset is regulated through a secure online portal, primarily available to verified researchers and industry professionals, with strict adherence to privacy and ethical use standards.</i>
<b>Data archiving and preservation</b>	<i>The dataset is securely archived in a cloud-based system with regular updates and backups, ensuring long-term data integrity and compliance with evolving technological standards.</i>
<b>Process for synthetic data creation</b>	<i>For the creation of synthetic data in this context, advanced simulation software is employed to generate realistic manufacturing environments and scenarios, using AI algorithms to model the behaviour and appearance of workers, safety equipment, and mobile robots. This process ensures a diverse and comprehensive representation of real-world conditions, enhancing the dataset's robustness for training and analytical purposes.</i>

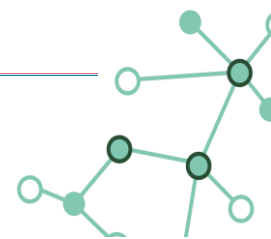




## 5.8 SAP Use Case: Data Privacy Preserving Analytics

---

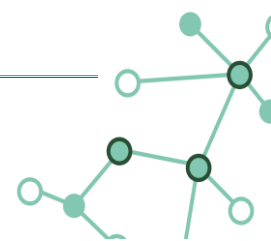
Name	Description
<b>Dataset Name</b>	<i>Exact data sets to be used are yet to be determined. See</i>
<b>Data provider and reference</b>	<i>We will use publicly available data gathered for research purposes, e.g., <a href="https://archive.ics.uci.edu/datasets">https://archive.ics.uci.edu/datasets</a></i>
<b>Purpose of Data Collection</b>	<i>The publicly available data will be used to train and analyse machine learning models and to test analytical tasks in the context of secure collaborative computations and our research prototypes.</i>
<b>Dataset Description</b>	<i>We will be using data sets suitable for the expected tasks with a focus on regression tasks in the context of, e.g., business applications and/or manufacturing insights.</i>
<b>Relation to the project</b>	<i>The data will be used to evaluate our research prototypes in the context of use case 3 and its associated protection mechanisms.</i>
<b>Standards and metadata</b>	<i>Some meta-data that are already provided by publicly available data are, e.g., subject area, task, number of instances, number of attributes and attribute type (see details about data provider)</i>
<b>Data sharing and access</b>	<i>The data is already publicly available and access regulations are up to the providers.</i>
<b>Data archiving and preservation</b>	<i>The data is publicly available, and its storage, security and preservation are up to the providers.</i>
<b>Process for synthetic data creation</b>	<i>As of now, there is no dedicated process for synthetic data generation as this Use Case intends to use the above mentioned publicly available datasets.</i>





## 5.9 WP8: GLACIATION Stakeholders Group

Name	Description
<b>Dataset Name</b>	<i>GLACIATION Stakeholder Group for Dissemination and Communication</i>
<b>Data provider and reference</b>	<i>This stakeholder dataset was gathered through voluntary participation via the GLACIATION project website, where individuals joined the GLACIATION community and consented to project-related communications. It represents engaged stakeholders interested in the project's goals and progress.</i>
<b>Purpose of Data Collection</b>	<i>The primary purpose of collecting contact information and identifiers from individuals signing up for the newsletter is for communication and dissemination activities related to the GLACIATION project. This includes updates on project progress, event participation opportunities, as well as general awareness about the project and its goals.</i>
<b>Dataset Description</b>	<i>The dataset comprises names and surnames of individuals who voluntarily sign up for the newsletter. It includes contact information such as email addresses, organisation details, and roles within the organisation. The dataset is limited to the necessary identifiers for effective communication.</i>
<b>Relation to the project</b>	<i>This dataset directly supports the communication and dissemination objectives of the GLACIATION project. It enables targeted outreach to stakeholders, fostering engagement and collaboration.</i>
<b>Standards and metadata</b>	<i>The collection and management of personal data adhere to the GDPR standards. Metadata includes information on the consent process, allowing individuals to opt-out at any time.</i>
<b>Data sharing and access</b>	<i>Access to the collected contact data is limited to its utilization within the GLACIATION project for communication and dissemination purposes, adhering to the consent obtained from stakeholders who voluntarily joined the GLACIATION community. This ensures that the use of data remains aligned with stakeholders' preferences.</i>
<b>Data archiving and preservation</b>	<i>The contact data is securely stored in a protected repository within Mailchimp, enabling efficient and agile communication practices while upholding the safeguarding of personal data in accordance with GDPR regulations. The secure storage measures will persist until the completion of the GLACIATION project or the individual opts-out, ensuring a reliable and compliant data archiving approach.</i>
<b>Process for synthetic data creation</b>	<i>Not applicable.</i>



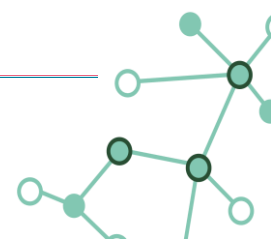


## 6 Data Protection Impact Assessment

The following section presents a DPIA for the stakeholder engagement dataset, as it contains personal data although of a non-sensitive nature. This assessment is essential to ensure that the use of these datasets poses no risk to individual rights and freedoms.

### 6.1 Stakeholder Engagement Dataset DPIA

Step 1: Identify the need for a DPIA
Please describe the objectives and the types of data processing involved?
<i>The objectives include communication and dissemination of information revolving around the GLACIATION project. Data processing activities involve the collection, storage, and use of personal contact information for newsletters and project updates.</i>
Please summarise the reasons for conducting a DPIA?
<i>To ensure compliance with GDPR, assess privacy risks, and implement necessary controls to protect personal data.</i>
Step 2: Describe the Processing
Please detail the methods for data collection, usage, storage, and deletion?
<i>Data is collected via voluntary sign-up on the project website. It is used for sending newsletters and updates. Data is stored in Mailchimp and will be deleted upon project completion or if the individual opts out.</i>
Please identify processing activities that are considered high risk?
<i>The potential for unauthorised access to personal data is considered a risk, requiring appropriate security measures.</i>
Please outline the data types, volume, frequency, retention period, number of individuals affected, and geographical area covered?
<i>Data types include names, emails, and organisational roles. The volume depends on sign-up rates, with frequency tied to project updates. Data is retained until project completion or opt-out and covers stakeholders globally.</i>
Please explain your relationship with the data subjects, their control over data, and their expectations?





*The relationship is based on voluntary consent to receive information. Data subjects have control over their data with the ability to opt-out at any time.*

Please state the intended outcomes, effects on individuals, and the benefits?

*The intended outcome is effective stakeholder engagement, with individuals benefiting from timely project updates and related articles of interest.*

**Step 3: Consultation Process**

Please detail the approach for consulting stakeholders and justifying any exclusions?

*Consultation via direct communication channels and feedback forms on the project website, excluding non-respondents after repeated outreach attempts.*

Please indicate which internal parties will be involved?

*The project communication team, data protection officer, and IT security team.*

Please mention any plans to consult security or other experts?

*The GLACIATION project has an external advisory board, which includes two specialists in technology and ethics who consistently assess the ethical dimensions of the project, including the protection of data rights.*

**Step 4: Assess Necessity and Proportionality**

Please propose measures to reduce identified risks, detailing the effect on risk and residual risk status

*Restricted access, access controls, and regular security audits to minimise risks, with ongoing monitoring to assess residual risk.*

Please describe what information will be provided to data subjects?

*Clear information on data collection, use, rights, and protection measures.*

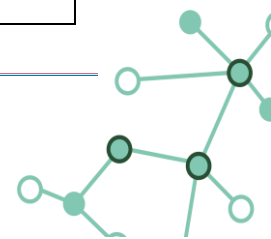
Please elaborate on how individual rights will be supported?

*Easy opt-out mechanisms and transparent data handling processes.*

**Step 5: Identify and Assess Risks**

Please describe each risk, its likelihood, severity, and overall risk rating.

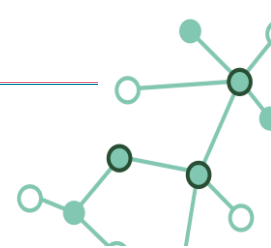
*Risks include unauthorised data access (medium likelihood, high severity) and inadvertent data sharing (low likelihood, medium severity).*







<b>Step 6: Identify Measures to Reduce Risk</b>		
Please propose measures to reduce identified risks, detailing the effect on risk and residual risk status?		
<i>Restricted access, access controls, and regular security audits to minimise risks, with ongoing monitoring to assess residual risk.</i>		
<b>Step 7: Sign off and Record Outcomes</b>		
Please record who has approved the measures and residual risks?		
<i>The project leader and DPO will approve the measures after a thorough review.</i>		
Please document any advice provided by the Data Protection Officer and actions taken?		
The current measures are stringent enough considering there is no sensitive data included. Beyond the measure already implemented regular monitoring and subsequent amendments and additional measures should the need arise.		
Please assign responsibility for the ongoing review of the DPIA		
<i>The DPO will be responsible for ongoing review, with annual updates to the DPIA.</i>		
<b>Name (DPO) and Date</b>	Alex Borg	20 <sup>th</sup> December 2023







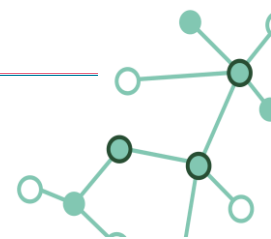
## 7 Annex GLACIATION Data Protection Impact Assessment Template

---

### GLACIATION Data Protection Template

This is the DPIA template for the GLACIATION project. All partners who handle datasets that include personal data should complete this document. It outlines the process of recording your DPIA process and results, in alignment with our DPIA guidance. Please consult this template in conjunction with that guidance and the Criteria for an acceptable DPIA as set out in European guidelines on DPIAs.

<b>Step 1: Identify the need for a DPIA</b>
Please describe the objectives and the types of data processing involved?
Please summarise the reasons for conducting a DPIA?
<b>Step 2: Describe the Processing</b>
Please detail the methods for data collection, usage, storage, and deletion?
Please identify processing activities that are considered high risk?
Please outline the data types, volume, frequency, retention period, number of individuals affected, and geographical area covered?





Please explain your relationship with the data subjects, their control over data, and their expectations?

Please state the intended outcomes, effects on individuals, and the benefits?

**Step 3: Consultation Process**

Please detail the approach for consulting stakeholders and justifying any exclusions?

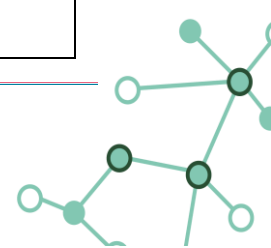
Please indicate which internal parties will be involved?

Please indicate which internal parties will be involved?

Please mention any plans to consult security or other experts?

**Step 4: Assess Necessity and Proportionality**

Please propose measures to reduce identified risks, detailing the effect on risk and residual risk status





Please describe what information will be provided to data subjects?

Please elaborate on how individual rights will be supported?

**Step 5: Identify and Assess Risks**

Please describe each risk, its likelihood, severity, and overall risk rating.

**Step 6: Identify Measures to Reduce Risk**

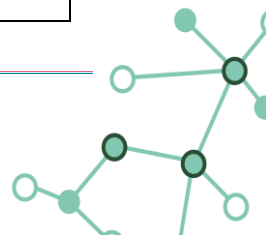
Please propose measures to reduce identified risks, detailing the effect on risk and residual risk status?

**Step 7: Sign off and Record Outcomes**

Please record who has approved the measures and residual risks?

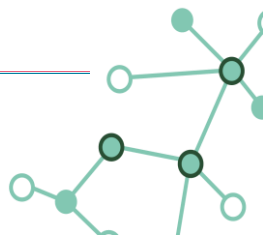
Please document any advice provided by the Data Protection Officer (DPO) and actions taken?

Please review and document responses from the consultation, explaining any deviations from stakeholder views.





Please assign responsibility for the ongoing review of the DPIA		
<b>Name (DPO) and Date</b>		





## 8 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](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](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

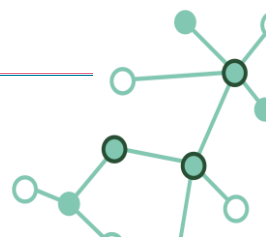
Regulation (EU) 2016/679 (General Data Protection Regulation).

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>





Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

