D1.2 Data Management Plan and Research Ethics (version 1)

Document Identification			
Status	Final	**Due Date**	28/02/2023
Version	1.0	**Submission Date**	17/03/2023

Related WP	WP1	**Document Reference**	D1.2
Related Deliverable(s)		**Dissemination Level**	PU
Lead Participants	KUL, DBC	**Lead Authors**	Theano Karanikioti, Peggy Valcke, Nikos Avgerinos
Contributors	ATOS, FSDE, FUJ_LU, INTRA, SQUAD, NORB, EXUS, LSTECH, SVI, QBE, SQD, XLAB, FN, ANYS, CEA, FHG, VTT, TUD, UMU, UTH, UPRC, IDSA, EGI, ECO, LIC, IDIADA, ABI, FMAKE, CESGA, AHOP, VISA, MET, UoM, UoG	**Reviewers**	Nikos Avgerinos (DBC) Claudia Mertinger, Jürgen Neises (FSDE)

Keywords:
Data Management Plan, Research Ethics, Data Protection Policy, Responsible Research and Innovation.

Document Information

List of Contributors	
Name	Partner
George Athanasiou	DBC
Tomás Pariente Lobo, Iván Zaldívar Santamaría	ATOS
Jürgen Neises, Claudia Mertinger	FSDE
Moussa Ouedraogo, Sofiane Lagraa	FUJ_LU
Panos Matzakos, Maria Fritzela, Eleni Veroni	INTRA
Pedro Pina, Pedro Cabral Santos, Marisa Brioso, Liliana Violante	SQUAD
Krishnapriya Sankaralingam, Vitalii Demianets	NORB
Konstantinos Kentrotis, Cristina Nichiforov	EXUS
Evangelos Kotsifakos, Pelayo Fernandez Blanco, Pavlos Kalfantis	LSTECH
Antonios Chronakis, Ioannis Spyropoulos, Pantelis Velanas	SVI
George Sachpatzidis, Niklas Palaghias, Thanassis Kountzeris	QBE
Elissavet Zogopoulou, Megaklis Vasilakis, Stavros Theocharis, Charis Giaralis, John Zaras, Nikos Ntampakis, Eftihis Thergiakis	SQD
Nejc Bat	XLAB
Chariton Palaiologk	FN
Dolores Ordóñez, Juan Ortells, Tayrne Butler	ANYS
Nicolas Belleville	CEA
Avikarsha Mandal	FHG
Ville Ollikainen, Anni Karinsalo, Outi-Marja Latvala, Visa Vallivaara, Sami Lehtonen, Aada Illikainen	VTT
Kaitai Liang, Zeshun Shi	TUD
Antonio Skarmeta, Jesus Garcia	UMU
Apostolos Apostolaras, Stavroula Maglavera	UTH
Asterios Stroumpoulis, Evangelia Kopanaki, Elena Avatangelou, Ioannis Katsanakis	UPRC
Giulia Giussani, Antoine Garnier	IDSA
Valeria Ardizzone	EGI
Lauresha Memeti, Nils Klute	ECO
Alessandro Paciaroni, Francesco Mureddu	LIC
Raul Villalba	IDIADA
Anzellotti Emiliano, Marco Crabu, Piperno Piero	ABI
Dries Verhees, Hugo Steep	FMAKE
Juan Carlos Cano, Tomeu Llobera Prats, Miquel Martorell, Jaume Ordinas	CESGA
Juana Maria Serra	AHOP
Andreas Kopysov, Vasili Schewelow	VISAR
Georgia Kyriakopoulou, Ioannis Vlachonikoleas, Antonis Giannopoulos, George Tsiatiris	MET
Athanasios Stratikopoulos, Christos Kotselidis	UoM
Sakshyam Panda, Emmanouil Panaousis	UoG

Document History			
Version	Date	Change editors	Changes
0.1	20/12/2022	Theano Karanikioti (KUL), Peggy Valcke (KUL), George Athanasiou (DBC), Nikos Avgerinos (DBC)	Table of Contents ("ToC")
0.2	29/12/2022	Theano Karanikioti (KUL)	Updated ToC with allocation of responsibilities
0.3	05/01/2023	Theano Karanikioti (KUL)	First draft of D1.2 (excluding Section 3 and Annex 2) for internal review within KUL
0.4	03/03/2023	Theano Karanikioti (KUL), Nikos Avgerinos (DBC), Claudia Mertinger (FSDE), Panos Matzakos (INTRA)	Second draft of D1.2 for internal review within KUL, including Section 3 and Annex 2 led and drafted by DBC
0.5	08/03/2023	Theano Karanikioti (KUL), Nikos Avgerinos (DBC)	Draft submitted for internal review
0.6	10/03/2023	Jürgen Neises (FSDE), Theano Karanikioti (KUL), Nikos Avgerinos (DBC)	Updated draft following internal review
0.7	15/03/2023	Theano Karanikioti (KUL)	Final formatting changes implemented
1.0	17/03/2023	Theano Karanikioti (KUL)	FINAL VERSION TO BE SUBMITTED

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Peggy Valcke (KUL)	15/03/2023
Quality manager	Jürgen Neises (FSDE)	16/03/2023
Project Coordinator	Tomás Pariente Lobo (ATOS)	17/03/2023

Table of Contents

List of Tables

List of Acronyms

Abbreviation / acronym	Description
AI	Artificial intelligence
API	Application programming interface
CA	Consortium agreement
DoA	Description of action
DMP	Data management plan
DPIA	Data protection impact assessment
DPO	Data protection officer
Dx.y	Deliverable number y, belonging to WP number x of the TANGO project
e.g.	Exempli gratia (latin phrase for “for example”)
EC	European Commission
EEA	European Economic Area
EOSC	European Open Science Cloud
EU	European Union
FAIR	Findable, accessible, interoperable and re-usable
GA	Grant agreement
GDPR	General Data Protection Regulation
HE	Horizon Europe
HE Regulation	Horizon Europe Regulation
i.e.	Id est (latin phrase for “that is”)
IDS	International Data Space
IPR	Intellectual property right
IT	Information technology
MFA	Multi-factor authentication
ML	Machine learning
R&I	Research and innovation
RRI	Responsible research and innovation
SDK	Software development kit
SME	Small and medium-sized enterprise
SSO	Single sign-on
TANGO DMP	The data management plan of the TANGO project
TANGO	Horizon Europe project TANGO (Digital Technologies ActiNg as a Gatekeeper to information and data flOws), Grant Agreement No.101070052
Tx.y	Task number y, belonging to WP number x of the TANGO project
VPC	Virtual private cloud
VPN	Virtual private network
WP	Work package
WP29	Article 29 Working Party

Executive Summary

This document constitutes deliverable D1.2 of the TANGO project, funded by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101070052. It sets out the first version of the data management plan and the research ethics and compliance protocol of the TANGO project and identifies relevant responsible research and innovation practices. It thus provides initial information about, among others, the data that is expected to be collected, generated and/or processed by the consortium within the remit of the TANGO project, and the manner in which such data and other research outputs are expected to be handled in terms of, *inter alia*, storage, security, and availability, while also laying out considerations related to making data and other research outputs findable, accessible, interoperable and re-usable. The document, furthermore, touches upon ethical and legal requirements, including but not limited to those related to data protection, that shall be considered and followed by the consortium, and practices that shall guide research activities during the project.

This deliverable thus relates to all work packages of the TANGO project, and all the activities undertaken by the consortium during the project shall be performed by reference to this document.

The information provided in this document is preliminary and non-exhaustive and will be refined during the project lifecycle. Activities undertaken under other work packages/tasks of the project are also expected to complement and/or influence the information presented in this document. An updated version of the matters covered in this deliverable will be provided in D1.3 – Data Management Plan and Research Ethics (final version) – due in month 32 of the TANGO project.

1 Introduction

1.1 Project overview

The TANGO (Digital Technologies ActiNg as a Gatekeeper to information and data flOws) project will create innovative solutions that will establish a stronger cross-sector data sharing in a citizen-centric, secure, trustworthy, and environmentally sustainable manner, promoting digitally enabled interactions across society, for people as well as for businesses. The overall TANGO environment will allow for user-friendly, secure, trustworthy, compliant, fair, transparent, accountable, and environmentally sustainable data management, having at its core technology components for distributed, privacy-preserving and environmentally sustainable data collection, processing, analysis, sharing and storage.

With partners comprising a balanced variety of renowned research institutes and universities, large, as well as small and medium-sized enterprises ("SMEs") that lead European research and development, and prestigious associations and think tanks, TANGO will provide a high impact solution within the transport, e-commerce, finance, public administration, tourism, and industrial domains. Through the provision of TANGO technologies, a trustworthy environment will be designed acting as a gatekeeper to information and data flows. Citizens and public/private organizations will be empowered to act and interact, providing data both online and offline.

More information about TANGO can be found on the dedicated website: www.tango-project.eu.

1.2 Purpose of the document

This document (deliverable D1.2) comprises the first version of the data management plan ("DMP") (led by KU Leuven – "KUL") and the research ethics and compliance protocol ("Protocol") of the TANGO project (led by Diadikasia Business Consulting – "DBC"). It also makes a preliminary identification of relevant responsible research and innovation ("RRI") practices (led by KUL). The document forms part of T1.4 "Data Management, Research Ethics and Legal Compliance", falling under work package ("WP") 1, that is, "Project and Technical Management".

Consequently, the purpose of this deliverable is three-fold: *First*, to describe the general categories of data 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, to a certain extent, after the end of this project (including, but not limited to, the processes used to gather the data, to secure the data and to make it available). An important part of the DMP concerns consideration of the so-called FAIR Guiding Principles – that is, how data (and other research outputs) will be made findable, accessible, interoperable, and re-usable ("FAIR"). *Second*, to establish the procedures that shall be followed by TANGO partners in order to ensure that data protection and ethical standards are met, and appropriate measures are taken throughout the research process. In doing so, it introduces, *inter alia*, the purpose, core concepts and general principles of and data subject rights enshrined in Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation or "GDPR" or "Regulation"), it presents proposed measures related to the data protection policy of the project and lays out data management guidelines, it provides guidelines about data protection impact assessments ("DPIA") and touches upon ethical issues and societal concerns in TANGO. *Third*, to identify, in a preliminary manner, RRI practices that will be relevant for the consortium's joint and individual research efforts during the project.

1.3 Relation to other project work

The DMP and the research ethics and compliance protocol are related to all WPs of the TANGO project. In other words, the activities in all WPs shall be performed in accordance with the considerations, principles, procedures, and guidelines laid out in the DMP and the Protocol. The same stands for the RRI practices identified, which are relevant for all research activities of the project.

Work carried out under other WPs/tasks will complement the DMP, the Protocol and/or the RRI practices identified in this document, and will add, further define and/or adjust requirements, processes and guidelines, and address matters related to data management, research practices, ethics, and compliance. Such tasks may include, indicatively:

- T2.2, which will extract user needs and requirements;
- T2.6, which will carry out a comprehensive Privacy, Ethical, Social and Legal Impact Assessment;
- T7.1, which will organize and prepare the pilot demonstrations to be executed during the TANGO project;
- T8.1, which relates to dissemination and communication activities;
- T8.2, carrying out market analysis and business modeling, including the preparation and collation of partners' exploitation plans;
- T8.3, which deals with intellectual property rights ("IPR") and innovation management;
- T8.4, which, among others, sets out the commercial roadmap of the TANGO solution;
- T8.5 relating to standardization activities; and
- WP9, addressing specific ethics issues.

In particular, work undertaken under WP2/T2.2 and/or WP7/T7.1 will, *inter alia*, lead to the determination and identification of the data that will be provided/used by use cases partners in the context of the project and that may, consequently, be made available to (certain) technical partners, as well as set the ground for collaboration between use cases and technical partners. It will also identify any interactions with external stakeholders as part of pilot demonstrations, thus relating to RRI practices to be followed during the project.

The Privacy, Ethical, Social and Legal Impact Assessment (WP2/T2.6) will, among others, identify the key legal, ethical, social and privacy themes of the TANGO platform, map data flows amongst TANGO technologies, users, and the services with which they interface, identify key risks stemming from these data flows and provide suggestions about any necessary technical or operational solutions and mitigation measures. As a result, this task will complement and is expected to specify arrangements that relate to this document. WP9 will also add to ethics themes of the project.

Finally, work carried out in WP8 will relate to the DMP and RRI practices preliminarily identified in this document, as it will, among others, contribute to establishing how data and research outputs of the project are to be handled and disseminated to the public.

1.4 Structure of the document

This document comprises five "substantive" sections. Following the introductory **Section 1**, **Section 2** lays out the first version of the TANGO DMP. The DMP refers to the data that will be handled by the TANGO partners (and how it will be handled), considerations regarding the "FAIRification" of data, and touches upon the management of research outputs other than data. It also discusses data management responsibilities under the project and the allocation of resources, and briefly refers to ethics considerations. **Section 3** comprises the research ethics and compliance protocol of the project. This section presents the purpose and core concepts of the GDPR, as well as general principles of data protection and rights of data subjects under this Regulation. It also introduces proposed measures regarding the project's data protection policy, and touches upon data management and measures, as well as DPIAs. The section also refers to ethical issues and societal concerns in TANGO. **Section 4** then includes an initial identification of RRI practices that should guide research activities under TANGO, such as open science, public engagement, ethics and integrity, and gender equality and inclusiveness. **Section 5**, finally, concludes.

The document also contains a section including the references used (**Section 6**), as well as two **Annexes. Annex 1** presents the questions asked to partners to collect initial information in preparation of the first version of the TANGO DMP. **Annex 2** presents the template data protection policy of the TANGO project.

1.5 Disclaimers

This deliverable constitutes the first version of the DMP, the research ethics and compliance protocol and the identification of relevant RRI practices, which will be refined during the project lifecycle. In other words, this version does not cover all the data management and research ethics matters in an exhaustive and final manner, such matters being subject to development, clarification, adjustment and agreement as the project evolves. It also does not exhaustively address RRI practices. A final version of the DMP, the research ethics and compliance protocol, and RRI practices will be submitted as deliverable D1.3 to the European Commission ("EC") in month 32 of the TANGO project. The information provided in the DMP, the Protocol, the RRI practices part and the Annexes does not constitute legal advice. Any user of this information uses it at their sole risk and liability.

2 TANGO Data Management Plan

Section 2 of D1.2 lays out the first version of the DMP of the TANGO project. DMPs are documents outlining all aspects of the research data lifecycle – in other words, they address data organization and curation, as well as data access, preservation, sharing and deletion. [1] Consequently, the TANGO DMP will set out how data collected, processed and/or generated during the project should be handled during its lifecycle and will refer to the procedures relevant to making data FAIR. As indicated above, the DMP is a living document, iteratively adjusted, updated, and enriched as the project evolves, taking into account, among others, the generation or use of new data, changes to the original planning, changes in data/output access provisions or curation policies, or changes in consortium policies, practices or composition. [1]

The TANGO DMP is based on, and to a large extent reflects, the Horizon Europe ("HE") Data Management Plan Template, Version 1.0, dated 5 May 2021. [2] Additional guidance documents, information and/or good practices have also been taken into consideration in preparation of the DMP, including EC documents (such as the HE Program Guide) and third-party resources. [1][3][4] Furthermore, requirements and obligations imposed by applicable – EU and/or national – legislation, including but not limited to data protection and privacy legislation shall guide data management during the TANGO project.

This section is structured as follows: **Sub-section 2.1** provides the data summary, in other words, it describes the data expected to be handled by TANGO partners in carrying out their research activities under the project. **Sub-section 2.2** touches upon data access, storage, and security during the project. **Sub-section 2.3** presents the FAIR Principles and preliminarily identifies how the consortium will take them into account, while **Sub-section 2.4** refers to the management of research outputs other than data. **Sub-section 2.5** describes the data management responsibilities and allocation of resources, and, finally, **Sub-section 2.6** briefly touches upon ethics issues.

Information presented in this section derives from TANGO partners' individual responses to the "TANGO Data Management Plan Questionnaire" circulated on 30 November 2022 (the questions are included in **Annex 1**), and further information provided by partners during the preparation of this deliverable. Partners that have contributed to the first version of the TANGO DMP include ATOS, FSDE, FUJ_LU, INTRA, SQUAD, NORB, DBC, EXUS, LSTECH, SVI, QBE, SQD, XLAB, FN, ANYS, CEA, FHG, VTT, TUD, UMU, UTH, UPRC, IDSA, EGI, ECO, LIC, IDIADA, ABI, FMAKE, CESGA, AHOP, VISA, MET, UoM, UoG.

2.1 Data summary

This sub-section provisionally outlines the data/information expected to be handled by TANGO partners during the project, describing the data that will most likely be collected, generated, re-used and/or processed in research activities. As indicated above, this is a first, high-level overview of such research data, as it is based on early-stage information collected from partners. Most partners have indicated that the data(sets) they will use have not yet been determined (or have not yet been defined with certainty), given that either they – or partners whose data they will use or whose needs and requirements are relevant for the determination of research activities – are still in planning and definition phase, or the respective tasks have not yet commenced. Consequently, answers to all questions of the questionnaire could not yet be provided or could not yet be provided in a definitive manner. Partners are expected to be able to provide further information as the project carries on. Such information will be included in the next iteration of the DMP.

In general, the TANGO project is expected to collect, generate and/or use the broad categories of research data/information shown in Table 1 below (the list and examples provided being preliminary and non-exhaustive at this stage, and consequently subject to confirmation or change):

Table 1: Categories of research data handled in TANGO

Data category	Data category content/examples	Expected data use
Contact, administrative, and other details of and data/information about TANGO partners, as well as relevant documents.	Data may include, among others, names; titles; organizations; email addresses; telephone numbers; membership, participation and contacts in committees, bodies and/or associations; affiliations; payment details. The data was/will be collected from TANGO partners during the project, as appropriate, and will encompass data of the partner itself and/or its representatives.	Data is collected and processed to allow communication and liaisons between TANGO partners in the course of the project and enable the organization and execution of project-related activities (e.g., but not limited to, organization of meetings, payments, other research activities, etc.).
Contact details of and data/information about external stakeholders.	This category may include, among others and as required in each case, names; email addresses; organizations and role in organizations; informed consent and opt-out forms; checkboxes about privacy and terms of service; signatures; data about participants in workshops. This category also encompasses data gathered from the TANGO website (e.g., related to usage statistics, registered users, etc.). The data will be collected, for example, by TANGO partners organizing and engaging in activities involving external stakeholders during the project.	This data will be used to plan and carry out communication and dissemination activities (e.g., stakeholders' engagement, newsletter, registration in events), workshops, events and other project/research activities.
Agendas, presentations, minutes, notes, signature lists, recordings and other documents and data from meetings, workshops, and events.	This category includes data generated as part of research activities of the project and events organized in the context of the project.	Such data will be used to, among others, enable and/or document research activities of the project.
Information from interviews, surveys, questionnaires, structured feedback, records, recommendations from workshops or other activities, and other information collected or generated from TANGO partners and external stakeholders through contacts, as well as (internal) documents and reports.	Research data in this category comprise a variety of information and materials (to be) collected or generated during TANGO research activities. Information is/will be collected from (representatives of) TANGO partners and/or external stakeholders through, *inter alia*, questionnaires, interviews, case studies and other activities during the project, and will be generated by TANGO partners as part of their research activities.	Such data will be used to carry out tasks under the project and draft relevant deliverables.

Data category	Data category content/examples	Expected data use
Materials and data(sets) used/generated in the context of the design and/or development of technologies, components and services, the training of artificial intelligence ("AI")/machine-learning ("ML")/federated learning models, the performance of data analytics, the drawing of inferences, and/or the testing, evaluation and validation of models, techniques, technologies, components, solutions and services and the overall TANGO framework.	This category may include, for example, information and data from/related to measurements undertaken during project activities, as well as other metrics (such as data related to operational parameters of a mobile device, e.g., performance metrics); user/customer/citizen data/information; data about preferences, transactions, interactions and usage; survey data; design data; processed quality information and analysis results; other (electronic) documents (e.g., certifying and containing data or representing policies). This category encompasses both existing data (e.g., data from surveys held in the past, other historical data, public research data), as well as data that will be collected/generated during the project. Such new data collection/generation may take place, for example, and depending on the specific case, synthetically; through sensors deployed in the context of pilot demonstrations; internal communication buses; smartphones or computers via applications/software development kits ("SDKs") integrated with TANGO applications; surveys; forms filled by users; dedicated platforms (e.g., used for manually entering data or for uploading documents); application programming interfaces ("APIs"); tools/instruments/software used for measurements; experimentation while designing and prototyping. This category may also encompass technical documents, guidelines, plans and reports generated as part of such processes, as well as other documentation.	Such data will be used, among others, for carrying out technical research activities of the project, and for testing, evaluating and validating results (e.g., through the pilot demonstrations to be executed).
Concepts, designs, and software.	This category encompasses, for example, concepts on software architecture and design (e.g.,	Such data will be used for carrying out technical research activities of

Data category	Data category content/examples	Expected data use
	functions, components, data flow, etc.), software artefacts, source code (e.g., in textual and/or binary format).	the project, including as part of the pilot demonstrations.
Bibliographical materials, literature, and similar publications.	Such materials may comprise, among others, reports, journal articles, books, websites, legislation, case law, existing technical data, etc. Publicly available materials will be collected as necessary and appropriate during the project through the respective source, e.g., (online) libraries, journals, websites, etc.	Such data will be obtained and used as background of and to support research activities undertaken as part of the TANGO project.

Following the presentation, in Table 1 above, of the broad categories of research data expected to be handled in TANGO, some further information is provided below regarding this data. As is the case with the identification of research data, the information provided below is preliminary and subject to change and/or further specification as the project progresses.

Types, formats and size of data

The above categories of research data are expected to be in a variety of types, including (as a preliminary, non-exhaustive estimate, and subject to clarification, addition and change, as the project progresses) textual, numeric (e.g., int, float), alphanumeric, binary, boolean, image, audio/video, geolocation, datetime, and a variety of formats, such as docx, csv, json, pdf, xls, npy, png, xml. Given that the project is still at an early stage, it is not possible to estimate with certainty the total size of the data that will be generated. With regards to the expected size of datasets of individual partners, the preliminary estimates provided by partners vary, depending on the partner's role and activities in the project, while some partners have also indicated that at this stage the expected size of their datasets is unknown. More information will be provided in the next version of the TANGO DMP.

Data origin and potential modifications

The project will generally both use newly collected/generated data and will re-use existing data. The decision as to the data to be used will be taken by partners bearing in mind their tasks and activities under the project. Relevant considerations include, for instance, whether a particular project activity necessitates the collection of new data or can rely on existing data, as well as whether datasets meeting the requirements TANGO partners seek for the performance of their research activities are available. Some partners have indicated that they will not re-use existing data or that they do not at this stage expect that they will re-use existing data, while others have indicated that they are planning on doing so. Some partners are considering, but have not yet determined, whether they will use existing data. Collection of new data throughout the project will be done as necessary for each specific task. Existing data used by TANGO partners may comprise their own historical data or may be third-party data (e.g., reports, technical data, etc.), although, as of this stage of the project, partners have indicated that they are still considering the data they will be using. Existing data will potentially be re-used for research activities, data model and algorithmic development and training, validation of the performance of TANGO technologies, and as background for project work. Existing technology offerings are also expected to be used. The period covered by each dataset varies – with some expected to cover (part of) the duration of the project, while others extending in the past, either for a specified time period (e.g., data collected in MM.YY) or for an undefined period (e.g., when it comes to journal articles, websites, etc.).

Manipulations to some data(sets) are expected to take place, depending on the activities to be carried out under each task, as well as the evolution of project activities and technical solutions developed. For instance, newly collected data may be added to a dataset, obsolete data may be removed, or data may be updated. Existing data may be modified by partners before being (re-)used in the TANGO project, if required for the purposes of carrying out research activities. Datasets may be anonymized or pseudonymized, e.g., to be made available to other partners for the execution of their tasks under the project. Manipulations may also be executed to constitute data to be inserted into a given system processable in terms of quality and quantity (e.g., by applying techniques such as missing values treatment, outlier treatment, variable transformations, etc.).

Information regarding personal data

While some TANGO partners will not process personal data (or it is not yet clear whether the datasets they will use will include personal data),[1] it is envisaged that certain partners will collect and process personal data during TANGO project activities. Personal data that will likely be processed during the project may concern, for example, the following categories of stakeholders (the broader categories and the personal data processed to be confirmed at a later stage):

- TANGO partners (including names, organizations, email addresses, signatures, etc.);
- External stakeholders participating in workshops, interviews or other events and activities held by TANGO partners in the course of the project, or being informed about project activities and results (e.g., names, organizations, email addresses, signatures, etc.); and
- Customers/users/citizens involved in pilots (for example, but not limited to, name, gender, date of birth/age, nationality, preferences, biometric data, behavioural data such as data about movement, device usage/information and transactions).

While the datasets to be used by each partner for project activities are still to be defined, it is expected that some personal data will be pseudonymized or anonymized, while other will not be (e.g., because it would not be possible to carry out the activity in question with anonymized or pseudonymized data). Further information will become available as the project progresses and is expected to be addressed in relevant tasks/deliverables.

Certain partners are expected to only engage in "limited" personal data processing (e.g., contact details of other TANGO partners), while others may process personal data to a larger extent (e.g., as part of the piloting activities of the TANGO project). Some TANGO partners anticipate processing special categories of and other sensitive personal data (e.g., facial biometric data for identity verification, data required for immigration processes). Appropriate consideration and due respect shall be given to the applicable legal framework and the legal and ethical guidance provided by DBC in the research ethics and compliance protocol of the project (**Section 3**). Given that guidance provided in the Protocol is expected to influence TANGO partners' data processing decisions and activities (e.g., but not limited to, with regards to the processing of certain personal data, such as personal data of vulnerable people or children), it is imperative that discussions take place among the respective partners before processing of personal data commences (e.g., as part of the pilot demonstrations). Based on the information currently available, personal data to be included in datasets handled by TANGO partners will generally not be

[1] Article 4(1) of the GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject')". It also states that "an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Thus, only anonymized data (after the point of anonymization) or information that does not refer to natural persons falls outside the scope of the GDPR. Whether a person is identifiable can only be assessed on a case-by-case basis. When data is pseudonymized, meaning that personally identifiable information, e.g., the individual's name, is substituted with a unique identifier not connected to their real-world identity, the GDPR still applies.

transferred to countries outside the EU/EEA,[2] although a partner has indicated that this may happen according to their organization's privacy policy and another partner has indicated that it is not yet known whether this will happen.

To the extent the processing of personal data is involved, it shall be ensured that it will take place in accordance with applicable EU, international and national law on data protection, notably the GDPR. [5][1] In particular, the data controller – that is, the natural or legal person who, "alone or jointly with others, determines the purposes and means of the processing of personal data" [5] – 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:

- Lawfulness, fairness and transparency;
- Purpose limitation;
- Data minimization;
- Accuracy;
- Storage limitation;
- Integrity and confidentiality; and
- Accountability.

In other words, it shall be ensured that personal data is:

- Processed lawfully, fairly and in a transparent manner;
- Collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;
- Adequate, relevant and limited to what is necessary in relation to the purposes for which it is processed;
- Accurate and, where necessary, kept up to date, with every reasonable step being taken to ensure that personal data that is inaccurate, having regard to the purposes for which it is processed, is erased or rectified without delay;
- 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
- 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. [5]

Finally, the data controller shall be responsible for and be able to demonstrate compliance with all the above. Further details as to the applicable legal requirements related to the processing of personal data and the data protection policy of the project are provided in **Section 3** of this deliverable.

Data utility

It is currently not possible to determine with certainty which of the research data generated and/or used as part of the project's research activities will be useful outside the TANGO project ("data utility"). This will be made clear as the project progresses. It is envisaged that certain research data collected, generated and/or processed during the TANGO project activities will not have, as such, a utility outside the project, being relevant and useful only for the execution of research activities during TANGO and for the preparation of the relevant deliverables. Other research data may have a utility outside the project, e.g., for researchers and academic community, for other EU-funded projects, for members of a particular industry/community, for third parties (private and public) pursuing solutions for similar projects as those tackled by the project's use cases, or to act as examples of the TANGO technology and document how it works.

[2] As has been flagged to partners, a data transfer does not entail actually "sending" the data to a non-EU/EEA country. If a partner or service provider is located outside the EU/EEA and is able to access the personal data collected, this also constitutes a "data transfer".

2.2 Data storage, access, and security

Following the data summary – which also touches upon data collection – in **Sub-section 2.1** above, this sub-section focuses on data storage, access, and security. Hence, **Sub-sections 2.1 and 2.2** cover the research data lifecycle.

While handling data, including storing and sharing data, the consortium shall consider and comply with requirements, obligations and standards set out in applicable legislation and guidelines, including – but not limited to – the GDPR. To the extent that the storage, sharing or other processing of personal data is involved, the fundamental data protection principles briefly outlined in the previous section should be respected. Further details about the legal and ethical standards and principles TANGO partners shall comply with are provided in **Section 3** of this document and will be further developed during the project.

2.2.1 Data storage, quality, and security

Broadly speaking (but without prejudice to specific circumstances arising from, e.g., the use of certain technologies/solutions and/or the nature/needs of specific activities in the context of TANGO, which may result in a different arrangement), data is expected to be stored by the TANGO partner(s) owning/providing each dataset. However, data storage may also be provided by the technical partner hosting the production environment used for pilots (although they will not provide datasets for the project). Certain data and information will also be stored on the project's common repository provided by the project coordinator. In addition, for some tasks/activities under the project, it is still not clear, and it is being discussed, whether to opt for a more distributed storage setting or a centralized one. Generally (though non-exclusively), it is anticipated that the partner(s) owning the dataset will control access to this dataset, and will be in charge of collecting, storing and deleting the data. However, it may be the case, for example, that a partner that is not responsible for collecting the data is responsible for deleting it (e.g., in the context of pilot demonstrations, where entities participating in pilots may be in charge of collecting data, with a TANGO partner being responsible for deleting it). Furthermore, data subjects will have control over their data, in line with applicable legal rules. These matters are, nevertheless, to be further assessed and confirmed later in the project, in the context of the activities to be carried out under each task/sub-task.

Various types of storage are expected to 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/Office 365, etc. – and/or be the project's common repository – OwnCloud – provided by the project coordinator). A combination of storage solutions may also be used depending on the specific research activities undertaken and/or a partner's choice, a decision which may also be influenced, among others, by the sensitivity of the data in question. It is possible, for example, that while the full dataset is stored locally, some limited data may be stored on the cloud, or that certain data/documents are stored locally, while examples and demonstrators are uploaded on the cloud, or that non-sensitive project data is stored on the cloud for collaborative work, while other data is stored locally. Source code will be stored on GitHub. Further decisions with regards to the storage are expected to be taken during the project. Partners also envisage to keep back-ups of research data, locally and/or on the cloud. Such back-ups are planned at appropriate intervals depending on the partner (ranging, e.g., from hourly to monthly). However, it should be noted that many partners have not yet defined the concrete plan regarding back-ups, and further information will be available as the project progresses.

Partners handling data shall also adopt appropriate measures to ensure data integrity, quality and confidentiality. In addition, data security is imperative, and TANGO partners shall protect data and information they hold by adopting necessary security measures and mitigating any risks.[3] Overall,

[3] Security is a fundamental principle with regards to personal data, and Article 5(1)(f) of the GDPR requires that appropriate technical or organizational measures shall be taken to ensure that personal data are protected "against unauthorized or unlawful processing and against accidental loss, destruction or damage". Thus, when processing personal data, TANGO partners shall have appropriate security measures and controls in place.

ensuring data confidentiality (i.e., that data is only available to those authorized and is protected against unwanted exposure and tampering), integrity (i.e., that data is protected and not altered to ensure that it is reliable, accurate and complete) and availability (that is, ensuring that authorized users can access the data in a timely and uninterrupted manner whenever necessary for carrying out activities under the TANGO project), in a balanced manner and in line with carrying out project activities is important. [6][7] Measures shall, therefore, be put in place, as necessary and appropriate, by TANGO partners to make sure that data access is restricted only to the intended audience (e.g., by adopting procedures for identification, authentication and authorization), as well as to prevent intentional or accidental destruction or modification of data (e.g., by maintaining back-ups and carrying out system audits).

Data integrity, quality, confidentiality, and security measures that have preliminarily been identified by TANGO partners and which may be adopted, as appropriate in each case, include:[4]

- Encryption at rest and encryption in transit methods/protocols;
- Integrity file system checks;
- Access controls (for example, implementation of appropriate restrictions to data access; role-based access control; appropriate measures for authentication and authorization, e.g., implementation of single sign-on ("SSO") with multi-factor authentication ("MFA"), use of SSH keys for authentication, etc.);
- Data only being accessible from the organization's cooperate network through the organization's laptops;
- Use of virtual private network ("VPN") to access data located on the organization's servers;
- Access to personal computers with password;
- Definition of conditions and policies under which data, research infrastructure and related tools and applications can be used, instructing users to follow set guidelines on how to use the tools and services that handle data, and enforcement of security, trust management and acceptable use policies;
- Use of cryptographic means for data generation (e.g., p-ABC signatures);
- Storage and/or sharing of documents that do not contain personal or confidential data;
- Local storage in hard drives with no internet access;
- Use of encrypted computer disks;
- Storage of data on the partner's infrastructure only with appropriate access control and restrictions;
- Firewalls to ensure network security;
- Use of tools with appropriate security measures;
- Protection of internal services communication by virtual private cloud ("VPC");
- Validating input data to ensure accuracy and veracity of information regarding recorded values;
- Replicating databases in production;
- Regular backups (which will also ensure data recovery);
- Periodic recovery tests to ensure recoverability;
- Access to back-ups only by information technology ("IT") services;
- Implementation of pre-processing techniques during data collection and at period sequence;
- Appropriate measures for physical security (e.g., physical access to premises for authorized persons only and through access control, guards in facilities);
- Regular security checks and audit controls;
- Opting for servers and services located in the EU to make sure that data handling is GDPR-compliant.

Furthermore, to ensure data integrity and quality, partners responsible for gathering information from other partners (e.g., use case providers) will be in frequent communication. In addition, confidentiality rules binding TANGO partners as per the Grant Agreement ("GA") and the Consortium Agreement ("CA") are relevant when it comes to data confidentiality. Finally, it should be noted that internal security policies and procedures put in place by certain TANGO partners also address such aspects, hence being relevant in the context of TANGO, and so are the security measures of the project's common

[4] These are an overview of measures identified by individual TANGO partners in their individual responses, and each partner envisages to take some of these measures according to their specific circumstances and needs.

repository provided by the project coordinator. Further information is expected to become available as the project progresses, including as part of project tasks related to privacy, security, and data management.

2.2.2 Data availability and sharing between TANGO partners

Certain datasets held by TANGO partners will not be made available to other TANGO partners. Such datasets may be, e.g., those containing sensitive personal data or those that would not be useful to other partners for their activities under the project. In fact, it is also possible that certain datasets may only be available to part of the partner's team that will manage the data (which will be ensured through strong access control), and logging mechanisms will provide appropriate reporting as to who is accessing the data or any abnormal behaviour. Although certain datasets may be inaccessible by partners other than their owner, results derived from their use may be made available to all TANGO partners, through the project's repository and/or through publications.

Nevertheless, data availability and sharing of some form between TANGO partners is expected to take place, to carry out research activities under the project. While at this stage of the project it cannot be conclusively specified which partners will have access to/use other partners' datasets and which datasets these will be, at least two broad "types" of data sharing are envisaged:

- Data sharing between use cases partners and (certain) technical partners; and
- Data sharing between partners carrying out similar or complementary activities under the project and/or being involved in the same tasks.

In addition, certain data/information will be accessible by the TANGO consortium (e.g., when all partners shall contribute data/information) through the project's repository.

In the first case, it is envisaged, for example, that use cases partners will make (a curated version of) their proprietary datasets available to (certain) technical partners for the purpose of carrying out project activities. Technical partners may assist use cases partners, e.g., with the organization and execution of the pilots, as well as with dataset management, including, but not limited to, preparing specific datasets for use in the TANGO project, adding persistent identifiers, implementing standard metadata access protocols and, if necessary, transforming data for compliance with the FAIR Principles. Technical partners may also utilize use cases providers' data to, e.g., carry out analyses, generate source code and models, develop procedures, test techniques, etc., which may then be used in the context of the project and/or support project activities by other partners (e.g., pilot demonstrations). For datasets made available by use cases partners to technical partners, it is generally anticipated that the use case partner owning the dataset will control access to the dataset, including access mechanisms and processes to be implemented. Use cases partners are also generally expected to control other aspects in relation to their dataset, though this is to be determined on a *per pilot* basis as the project progresses (e.g. while some partners may provide anonymized data to the project, others may require technical partners' involvement in this). Matters relating to access to/use of use cases partners' datasets by technical partners also depend on the technology used in a specific case and the specific activities concerned. It has also been suggested that the potential use of data spaces (e.g., data spaces based on International Data Space ("IDS") standards) in the project could also address some matters regarding access to/usage of datasets during the project, as data providers can have control over what they share and with whom. This is also to be further assessed as the project progresses.

With regards to data sharing, to the extent necessary and appropriate, between partners having similar or complementary tasks under the project, this may take place to allow benefiting from synergies, to carry out technical or other work under the project and/or to utilize project resources more effectively (e.g., to avoid duplicating efforts of the consortium).

Broadly speaking, it is expected that it is partners requiring access to another partner's data to carry out defined activities under the project that will be given access to that data. However, certain non-sensitive data may be available to all TANGO partners, e.g., through the project's common repository (OwnCloud) provided by the project coordinator, accessible by TANGO partners. Partners shall ensure that data sharing will take place in accordance with legal requirements. Consequently, any necessary

conditions for or restrictions to access to or use of the data (e.g., authorization) shall be set out, as appropriate, which is generally anticipated to be done by the partner owning the dataset. Information about access to data shall also be provided to ensure that partners can undertake their research activities under the project. Furthermore, partners shall collaborate, as necessary and appropriate, to arrange aspects surrounding data access and sharing between them for the needs of the project (e.g., to come to an agreement with regards to data specifications, such as formats, restrictions, etc., to be used). In addition, partners shall adopt measures to ensure that data to be used by other partners during the project will be available to the latter when needed. Anonymization or pseudonymization of datasets may be required and/or performed, though, as of this stage in the project, and given that the datasets to be used are still being defined, it is not possible to ascertain which data made available between partners will be pseudonymized or anonymized (though it is anticipated that the sharing of original datasets will not always be required, given that certain project activities by, e.g., technical partners can be carried out with pseudonymized or anonymized datasets). This matter will be duly considered and decided upon in the following months of the project. Finally, combination of datasets is not yet concretely planned, but this matter is to be assessed, addressed, and determined by partners in the course of the project (also in light of applicable legal rules). Some partners have, however, already indicated that their datasets will not be combined with other datasets and/or that, if possible, datasets shall be separated.

At this stage of the project, a mapping of data availability and sharing between TANGO partners, as well as details about conditions for and any measures taken to ensure access to data are not yet possible. With regards to data availability and sharing between use cases partners and technical partners, work is currently undertaken as part of WP2/T2.2, which will, among others, identify the datasets that will be provided by use cases partners and set the ground for the cooperation between use cases and technical partners. When it comes to synergies between partners, this will also be clarified as the project carries on. It should also be noted that TANGO partners are expected to consider the issue of availability of data to other members of the consortium throughout the duration of their activities and to adapt their approach, e.g., by restricting access to certain data, if necessary.

2.2.3 Archiving, preservation, and deletion of data

The matter of archiving, preservation and deletion of data is to be revisited and further addressed at later stage of the project, and related information will be provided in D1.3. Applicable organizational, legal, and regulatory requirements, rules and obligations will be taken into account in determining the approach to such matters.

On a preliminary note, it is anticipated that data will be stored until it is clear that they will not be analysed again for project activities and/or until the project ends and the final review has been undertaken. Data will subsequently be deleted and/or discarded from the storage that has been used during the project. Certain data and research results may, however, be kept (for a certain time) and/or archived after the project, while taking necessary and appropriate measures (e.g., blurring of personal data before archiving). This is to be further assessed later in the project. Datasets published/shared with third parties (e.g., if open access is provided) and published results will be kept after the project, and source code that has been generated during the project by certain partners may remain open-source on GitHub, while respecting the IPR strategy of the project. In addition, examples used to demonstrate technology, e.g., in the project's website or project deliverables, are expected to remain public.

2.3 Making data FAIR

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 TANGO project will take them into account. In addition, **Sub-section 4.1** below addresses open science, including open access, practices that will be considered by TANGO partners.

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). Therefore, 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. [8] The FAIR Guiding Principles comprise four elements, which are related, but independent and separable: findability, accessibility, interoperability, and re-usability. These principles are meant to be followed "in any combination and incrementally", considering the context and special circumstances of each case. These principles can be applied not only to data, but also to non-data assets. [8]

Responsible research data management in line with the FAIR Principles is a mandatory open science practice for HE-funded projects, notably through the use of DMPs and open access to research data under the principle "as open as possible, as closed as necessary". [1] An important observation should be made at this stage: FAIR data does not equal open data (that is, data that is publicly available for everyone to access and re-use). Data can – and shall be – FAIR even if access is restricted. [1] Such restrictions may be due to necessary and legitimate reasons (e.g., protection of personal data, protection of IPRs, trade secrets, etc.).

The following sub-sections provide some initial information about individual TANGO partners' considerations with regards to each of the four elements of the FAIR principles. However, given the initial stage of the project, (elaborate) information on this matter cannot yet be provided, hence data "FAIRification" is envisaged to be revisited. Specific arrangements with regards to steps to be taken in line with these principles will be made as the project progresses and be elaborated upon in D1.3.

2.3.1 Making data findable, including provisions for metadata

The first element of the FAIR Guiding Principles is "findability". 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". [8] Assigning a globally unique and persistent identifier – that is, an identifier that cannot be re-used/assigned without referring to the specific data for which it was initially assigned, and that is not rendered invalid over time – is of utmost importance in achieving other aspects of FAIRness. [9] Having rich metadata, including descriptive information as to the context, quality, condition and/or characteristics of the data, allows researchers and, importantly, computers to find data based on the information provided by its metadata, thus facilitating re-use. [10] Finally, registering or indexing the (meta)data – e.g., in repositories or specialized engines – ensures that it can be discovered on the internet and, consequently, that it can be re-used, as others can be made aware of the data's existence. [11]

TANGO partners will consider measures related to the findability of their data, as appropriate and taking into account the specific circumstances of each case, although at this stage of the project the concrete measures to be taken towards this goal cannot be identified with certainty. Measures that will be considered, and potentially taken, as appropriate, include assigning a persistent identifier to data and/or metadata, providing rich metadata (e.g., in terms of type of data, timestamps, etc.) and search keywords in the metadata to optimize the possibility for discovery and potential re-use. In addition, it is anticipated that specified naming conventions will be followed, version numbers will be provided, and metadata will be offered in a way that can be harvested and indexed. The use of trusted open access data repositories (e.g., Zenodo) will also be considered.

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 standardized communications protocol"; which (1.1) "is open, free, and universally implementable", and (1.2) "allows for an authentication and authorization procedure, where necessary"; and (2) "metadata are accessible, even when the data are no longer available". [8] In

other words, FAIR data access entails that access is mediated without specialized or proprietary tools or communication methods (e.g., protocols that have limited implementations or poor documentation) – notwithstanding the possibility that the access protocol is not fully mechanized, in case of, e.g., highly sensitive data. [12] Instead, the protocol used should be free and allow anyone with a computer and an internet connection to potentially access at least the metadata. [13] As explained above, however, FAIR data are not open data: hence, what is envisaged under the "accessibility" principle is that the exact conditions under which the data is accessible should be provided – in other words, ideally, it should be possible for a machine to automatically understand the requirements for access and then either automatically execute the requirements or alert the user about the requirements (e.g., the need to create a user account to authenticate). [14]

Taking the above considerations into account, TANGO partners will make an assessment, at a later stage during the project, as to the data to be made accessible to third parties, including the data to be made openly available. Based on that assessment (and in line with any applicable rules), partners will also consider the conditions of access to such data and how to appropriately specify such conditions. As of this moment, it is not yet possible to give concrete information on the steps to be taken towards accessibility of research data generated during the TANGO project. However, some general considerations are presented in this deliverable based on initial individual responses of partners to the TANGO Data Management Plan Questionnaire. The issue will be further addressed in the final version of the TANGO DMP.

Certain data may be included as examples, e.g., in technical documentations of solutions developed through TANGO. It is anticipated that this data will be made accessible accompanying such documentation (e.g., in a code repository, the project's deliverables, etc.). It is also anticipated that some research data will not be accessible to third parties as such. However, processed information, resulting activities, analyses, models, or other results are expected to be made available in, e.g., deliverables disseminated to the public, reports, journal articles, conference proceedings and other technical documentation of the TANGO solutions. Certain results may also be made available through open pre-print repositories (such as, hal and eprint.iacr). Code may be hosted on GitHub in a repository that supports read access by any user, certain technologies developed are expected to be available as open-source on GitHub and certain data may be made accessible through an Open API exposure. Research data underpinning scientific publications may be made available to third parties – under the premise "as open as possible, as closed as necessary" – at the time of the publication, in which case it is expected that it will be deposited in public repositories (e.g., gitlab or GitHub) or in the repository associated with the publication. To the extent shared, such data will generally be in a widely used format. In some cases, certain technologies, methods, or software tools may be needed to access the data. If available, it is considered that standardized protocols will be followed. Finally, metadata may be described in relevant publications that present project results and, to the extent that the data is published alongside the publication, metadata may also contain information to enable the user to access the data. Metadata reported in publications is expected to remain accessible even if published research data is no longer available.

2.3.3 Making data interoperable

For data to be re-used, it usually needs to be integrated with other data, and to interoperate with applications or workflows for analysis storage and processing. [15] The "interoperability" principle entails that: [8] (1) "(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation" – which would allow humans to exchange and interpret each other's data, and machines to read the data without the need for specialized or ad hoc algorithms, translators or mappings (which means that they should know the other system's data exchange formats); [16] (2) "(meta)data use vocabularies that follow FAIR principles" – meaning that the vocabulary used in describing the dataset shall be documented and resolvable using unique and persistent identifiers, with the information being easy to find and accessible by those who use the dataset; [17] and (3) "(meta)data include qualified references (i.e., cross-references explaining their intent) to other (meta)data" – in other words, it makes and sufficiently describes meaningful scientific links between (meta)data resources,

enriching the contextual knowledge about the data (e.g., specifying that one dataset builds on another, stating that complementary information is found in another dataset). [18] Thus, this principle refers to technical interoperability.

Broadly speaking, TANGO partners will seek to make data with a utility outside the project interoperable. Data held by TANGO partners is expected to generally use widely-known and community-endorsed vocabularies, standards, formats or methodologies (e.g., json, xml, ngis-ld, schema.org, W3C's Verifiable Credentials, csv files, DDI, Dublin Core), and necessary information to process the data and understand what it is by associating a publication with its metadata may be provided. Certain partners may also store data in an open data format, which can be opened with any programming language, and while making available the format specifications. In addition, specific ontology mappings are expected to be provided, as necessary. To the extent relevant, qualified references to other data may be included. Finally, it is likely that IDS-compliant data spaces may be used in the context of TANGO, which may address some data interoperability aspects, but this will be assessed further at later stages of the project.

2.3.4 Increasing data re-use

Increasing and optimizing data re-use is the ultimate goal of the FAIR principles. [15] To ensure that data is re-usable, "(meta)data [shall be] richly described with a plurality of accurate and relevant attributes. [8] This means that adequate and appropriate labels attached to the data should allow the (re)-user (human or machine) to decide if the data will indeed be useful for them in the particular context (by considering, for example, the context under which the data was originally generated). For this to happen, aspects such as the purpose for which the data was generated/collected, the date of generation/collection, who prepared the data, the software used, etc. should be described. [19] Moreover, (meta)data need to be "released with a clear and accessible data usage license", indicating clearly the conditions under which the data can be re-used; and it shall be "associated with detailed provenance" – i.e., it should be clear where the data comes from, who generated or collected it, how it has been processed, whether it has been published before and whether it contains data from other sources. Finally, it shall "meet domain-relevant community standards", as having data that is of the same type, that is organized in a standardized manner, that uses well-established and sustainable file formats, that is accompanied by documentation that follows a common template and that uses a common vocabulary makes it easier to re-use datasets. [8]

TANGO partners shall take a number of steps to increase the re-usability of their research data that may have a utility outside the project. Such steps will further be specified as the project progresses and may include, as necessary and appropriate, the following:

- Thoroughly documenting the provenance of data using appropriate standards;
- Documenting software and providing Git instructions;
- Including descriptions of the data in shared documents and describing formats, designs, and applications in technical documentations (e.g., methodology, fields used, purpose, etc.);
- Making available readme-files alongside the data, as well as notebooks to illustrate the use of data, instructions and guidelines;
- Making available through publications results of analysis carried out on the data, which will allow to validate further data analysis by comparing outcomes;
- Publishing examples alongside technical descriptions of the solution developed through public deliverables, repositories and/or research publications.

In addition, the use of appropriate common types of data sharing/re-use licenses will be considered (e.g., Creative Commons), but this is to be defined during the project. Quality assurance procedures will also be defined in the course of the project, and, among others, may include, e.g., downloading published data to check that it corresponds to data analyzed to generate the results (verification) or performing risk assessment for tools developed following applicable frameworks.

Overall, in determining the steps to be taken to allow for the re-use of data (and, more broadly, to define the measures to make data FAIR), appropriate consideration will also be given to IPR and innovation

management, addressed under WP8/T8.3 of the TANGO project. It should also be noted that even if the underlying data is not published, data models are expected to be re-usable (based on appropriate licenses), and so are results of analyses carried out to the extent published in publications.

2.4 Management of other research outputs

The TANGO project will produce digital research outputs other than data, including models, workflows, software, algorithms, frameworks, protocols, techniques, APIs and other interfaces, and services. Limited physical research outputs (printed samples) are also expected to be created. The project will also generate deliverables and other digital documents such as roadmaps, recommendations, white papers, and scientific publications.

Research outputs other than data will be managed having regard to the FAIR Guiding Principles, as relevant and appropriate, with the aim of making research available as openly as possible. As the project is still at an early stage, details as to the management of such research outputs will be clarified/decided later in the project. Accordingly, in this sub-section, a high-level initial overview is provided, while more detailed information will be included in the final version of the TANGO DMP (deliverable D1.3).

Documents produced within the scope of the TANGO project will follow standardized naming conventions, as specified in the Project Management Handbook (deliverable D1.1). [20] In particular, deliverables shall follow a set nomenclature – namely *Project_Dx.y_Name_vm.n_[suffix]* – which corresponds to the project short name, the deliverable number as defined in the description of action ("DoA"), the name of the deliverable, the version number of the document, and optionally a suffix used to identify intermediate versions or contributions from partners to a draft version. Deliverables shall also include a "Document Identification" part, where, among others, the version number shall be indicated, as well as include keywords and abbreviations. The above will facilitate the findability of the TANGO deliverables (most of which will be made openly accessible through the TANGO website), as well as navigating through them.

Furthermore, TANGO partners will make (peer reviewed) scientific publications openly available, using appropriate licenses and/or public repositories (e.g., Zenodo). Further information as to open access considerations is provided in **Sub-section 4.1** below, and details will be considered and defined during the project. Information regarding dissemination of certain research outputs (e.g., deliverables and publications) can also be found in deliverable D8.1. Scientific publications and deliverables made publicly available will be provided in widely used text formats, e.g., pdf.

Software will be managed according to the project's access policies to be defined at a later stage. It is expected that certain software/techniques will be open access, e.g., as code repositories (Git), open-source on GitHub, while other software (components) will not be open access and, e.g., be commercialized by TANGO partners. Other project outputs, like workflows or architecture, are also expected to be made public through, for example, public deliverables, and so are models under the appropriate licenses. Instructions, guidelines, and other files necessary to allow re-usability of a model may also be provided, as necessary.

Overall, the management of research outputs other than data created during the TANGO project will take into account the FAIR Guiding Principles, while also considering the confidentiality of the information disclosed by partners during the project and ownership of outcomes stemming from project activities, the planned commercial utilization of results, and the protection of IPRs (including patents), know-how and information related to the use of knowledge owned by a partner as a result of work carried out prior to the project. Further information on this matter will be provided as part of the activities of T8.3. More broadly, additional information about the management of research outputs other than data, including, but not limited to, publications, software components, models, and workflows, will be provided in deliverable D1.3.

2.5 Allocation of resources

Within the TANGO project, KUL has been appointed as the Data Manager, administering the processes of the construction of the TANGO DMP and data management more broadly. In addition, certain partners have identified, within their organization, employees responsible for data management for the purposes of the TANGO project. The construction of the DMP (first and final version), as well as data management within the remit of TANGO is a collaborative process, requiring (iterative) input and contribution of all TANGO partners. To this end, the consortium has been asked to provide information necessary for the drafting of the first version of the TANGO DMP, to review working drafts and to comment and contribute to them. Such an approach will continue to be followed in the course of the project, and TANGO partners have been invited by KUL to provide, on their own initiative at any stage during the project, information relevant to data management.

With regards to the handling of research data under TANGO, to the extent that a consortium partner collects and/or generates the data, that particular partner will be responsible for proper collection, storage, processing and sharing of that data, and for ensuring that personal data is treated in accordance with the applicable legal framework, unless specified differently by data processing agreements concluded between partners in line with the applicable legal framework.

At this stage, the costs partners will incur for making data and/or other research outputs FAIR cannot be established with certainty. Relevant costs to be incurred are expected to include, e.g., personnel costs, storage costs, and dissemination costs, including publication fees for providing open access to research outputs (e.g., scientific publications). Further information with regards to the costs incurred will be made available during the project and reported in D1.3 (final version of the Data Management Plan and Research Ethics). Costs incurred for providing open access to scientific publications are expected to be (partially) covered by the project funding each partner receives. Appropriate acknowledgment of the funding received from the EU will be provided. Some project-related costs may also be covered by the internal budget of TANGO partners.

Finally, while it is still to be defined by most TANGO partners how they will ensure long-term preservation of data (to the extent this is appropriate in the specific case), it has been suggested that back-ups, the dissemination of results in publications, the publication of pre-prints in open archives and the use of public, free-of-charge repositories could ensure the availability of published data and/or results for a long period of time.

2.6 Ethics

TANGO partners shall carry out their project activities in accordance with the highest ethical standards and applicable EU, international and national law. As required under Regulation (EU) 2021/695 establishing Horizon Europe – the Framework Programme for Research and Innovation ("HE Regulation"), appropriate attention shall be paid to, among others, the right to privacy, the right to data protection, the right to the physical and mental integrity of the person, the principle of proportionality and the need to ensure protection of the environment. [21] Any collection and processing of personal data of data subjects shall be done in compliance with applicable EU, international and national law on data protection, particularly the GDPR. Respect for the fundamental principles of the processing of personal data enshrined in Article 5 of the GDPR shall be ensured.

To make sure that ethical considerations are appropriately taken into account throughout the project and that ethics principles are properly identified and respected by the consortium, a dedicated partner has been appointed for continuous ethics monitoring (DBC). The first version of the research ethics and compliance protocol of the TANGO project has been defined and is presented in **Section 3** of this document. This Protocol sets out the ethical procedures that the consortium shall adhere to in order to ensure that adequate ethical standards are met, and data protection measures are taken throughout the research process. All partners are expected to review the Protocol and comply with it. Furthermore, a comprehensive impact assessment, covering ethical, social, legal and privacy issues will be carried out

under WP2/T2.6, which will further define and monitor legal and ethical requirements. In addition, specific ethics monitoring will be carried out under the dedicated "ethics" WP9 of the TANGO project.

3 TANGO Research Ethics and Compliance Protocol

In this section, the main concepts, and the most important Articles of the GDPR are thoroughly analyzed, while the basic principles that rule the processing of personal data as well as the main rights that the GDPR provides to the data subjects are also being mentioned. Furthermore, in this section the legal bases under which data is processed, the privacy policy which may be implemented by data controllers and the conditions for appointing a data protection officer ("DPO") are presented. Special reference has been made to the manner in which the security of processing is ensured in the TANGO project, but also to the manner in which we deal with a personal data breach situation. Moreover, the definition of the DPIA has been analysed and the cases in which it should be performed in the TANGO project have been mentioned. Finally, an ethical evaluation is being made with references to the fundamental principles of research integrity and cases in which ethical issues may arise during the processing of personal data.

3.1 The purpose of the GDPR and its core concepts

The purpose of the GDPR is to protect fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data, while also ensuring the "free movement of personal data". The GDPR sets out the EU regulatory framework for the processing of personal data. According to Article 3, the Regulation applies to the processing of personal data in the context of activities of an establishment of a controller or processor in the EU; to the processing of personal data of data subjects who are in the EU by a controller or processor not established in the EU, where the processing activities are related to offering goods or services to such data subjects in the EU or monitoring their behaviour as far as it takes place within the EU; as well as to personal data processing by a controller not established in the EU but in a place where EU Member State law applies on the basis of public international law.

In accordance with Article 1 of the GDPR, the Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. The Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

For a more effective understanding of the Regulation, the following definitions are considered indispensable, provided in Article 4 of the GDPR:

"Personal data" means any information relating to an identified or identifiable natural person ("data subject"). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

"Processing" means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction.

The processing may involve special categories of personal data. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. All this data is sensitive and belongs to special category of personal data. One of the cases

when processing of sensitive data is allowed is when the data subject has given explicit consent to the processing of this specific personal data for one or more specified purposes.

"Data processor" is defined in the GDPR as the natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller.

"Data controller" is defined in the GDPR [5] as the natural or legal person that, alone or jointly with others, determines the purposes and means of the processing of personal data. It is possible that a controller may involve another actor, known as a data processor, to process personal data on behalf of the controller and in accordance with the ways and purposes specified by the latter. This difference can be difficult to translate into the complexity of modern relationships but what is decisive is the scope of decision-taking power – i.e., who decides the "why" and "how" personal data shall be processed. Data controllers have direct obligations by handling data of data subjects, but data processors will be under direct obligations as well, such as the obligation to maintain a written record of processing activities carried out or a duty to notify the controller on becoming aware of personal data breach without undue delay. Data controllers must provide transparent information to data subjects at the time personal data is obtained, in a clear, comprehensive, and easily accessible way. The data subject must be informed about his/her rights, the way data is going to be processed, for which reason, but also, for example, about the period for which data is going to be stored.

Although the status of each partner (data controller or data processor) is not clear at the moment since we are still in the beginning of the project and many partners still do not have a clear knowledge regarding their duties, it is expected, during the progress of the TANGO project, that they will process personal data according to GDPR, such as ID and contact information, information regarding their work or marital status or any other data that could help identify a person, directly or indirectly. Moreover, some partners are also expected to process special categories of personal data (according to Article 9 of GDPR), such as biometric data for example.

As a result, the partners of the project should be in a position to recognize whether they are expected to process personal data and thus, need to respond properly to all the preliminary actions being taken to assure data safety by the authorized partners. Also, those who are due to engage in data processing activities, should respect the GDPR provisions, having into consideration the basic principles of the Regulation and the rights and freedoms of the data subjects.

3.2 General principles of data protection and rights of the data subjects under the GDPR

The basic principles that are required to be observed when processing personal data by data controllers and processors are set out below. The legal bases for the processing of personal data and the main rights of data subjects when processing their data are also extensively described. Those are considered to be the cornerstone of data safety and the partners of the project are obligated to respect them throughout their data processing activities.

The basic principles that are required to be observed during the processing of personal data from data controllers and processors are, according to Article 5 of the Regulation, the following:

- **The principle of lawfulness, transparency, and fairness of the processing of personal data.** Personal data must be processed in a lawful, transparent, and fair manner. The lawfulness of processing is ensured, in accordance with Article 6 of the Regulation, in cases where the prior consent of the data subject to the processing of his/her data for one or more specified purposes has been obtained, the processing is necessary for the performance of a contract or for compliance with a legal obligation of the controller arising from another rule of law, the processing is necessary to safeguard a vital interest, and finally processing is necessary for the purposes of the legitimate interests pursued by the data controller, including but not limited.
  Transparency is ensured by providing to the data subject all information on the processing in a concise, transparent, and comprehensible manner. Fairness means that processing must be done in

ways that data subjects would reasonably expect and not in ways that have unjustified adverse effects on them.

- **The principle of purpose limitation.** Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.
- **The principle of data minimization.** Personal data shall be adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed.
- **The principle of accuracy of personal data.** Personal data shall be accurate and, where necessary, kept up to date. Every reasonable step must be taken to ensure that personal data that is inaccurate, having regard to the purposes for which it is processed, is erased, or rectified without delay.
- **The principle of storage limitation.** Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed.
- **The principle of integrity and confidentiality.** Personal data must be processed in a way that guarantees appropriate security of personal data, including its protection from non-authorized or unlawful processing and accidental loss, destruction, or damage by using appropriate technical or organizational measures.
- **The principle of accountability,** under which the controller and processors are responsible for, and must be able to demonstrate compliance with, the principles relating to the processing of personal data.
- **The principle of proportionality**, which requires that there must be a connection between data kept and the purpose for which it is collected.

According to Article 6 of the Regulation, in order for the processing of personal data to be lawful, it must be based on a legal basis. The GDPR requires any organization processing personal data to have a valid legal basis for that processing activity. The GDPR provides six legal bases for processing personal data: **consent of data subject**, **performance of a contract** (the data processing should be lawful where it is necessary in the context of a contract or the intention to award a contract), **a legitimate interest pursued by the data controller or by a third party** (the legitimate interest assessment is carried out in three stages. First, the actual legitimate interest of the controller is assessed. Secondly, it is assessed whether the envisaged processing is strictly necessary for its fulfilment and thirdly, whether the processing overrides the interest of the data subject not to have his/her data processed for reasons of privacy), **a vital interest of the data subject or of another natural person** (processing is necessary in order to protect the vital interests of the data subject or of another natural person), **a legal requirement** (the processing is necessary for compliance with a legal obligation of the data controller), and **a public interest** (processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested on the data controller).

Article 7 of the Regulation prescribes conditions for demonstration of consent. “Consent” of the data subject means any freely given, specific, informed, and unambiguous indication of the data subject’s wishes by which he/she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her. When consent is given in a written form, it should be clearly distinguishable from the rest of the information provided to the data subject. The informed consent shall leave no possible space for ambiguity or confusion for the data subject, and it should make clear that the aim is to collect and process their personal data. The consent should use clear and plain language, which means that its content shall not be hidden behind complex legal formulas. It is necessary to inform the data subject of the possibility to withdraw their consent at all times and without need for any uneasy procedures. Consent to process sensitive data will have to be explicitly given by a data subject and they must be given the capability to withdraw consent effortlessly. This is confirmed by Article 7(3) of the Regulation, which states that the data subject shall have the right to withdraw his or her consent at any time and it shall be as easy to withdraw as to give consent. When the data subject has given explicit consent to the processing of sensitive personal data for one or more specific purposes, then the processing of such personal data is permitted (see in **Sub-section 3.1** the definition of special category of personal data). Existing consent given under previous rules may still be valid, but only given that they meet the new stricter requirements. The withdrawal of consent shall however not affect the

lawfulness of processing of data based on consent given before its withdrawal and the data subject shall be informed thereof.

Before personal data is collected from the data subject, the data controller shall provide him/her with a variety of information prescribed in Articles 13 and 14 of the GDPR, such as the identity of the controller, contact details, the purpose of the processing of data (lawful processing), the period for which data is going to be stored, the existence of the right to request from the controller access to and rectification or erasure of personal data, the right to lodge a complaint with a supervisory authority, etc.

The GDPR further acknowledges a range of rights for data subjects whilst the existence of these rights should be brought to the attention of the data subject in the explicit and clear manner in the informed consent. Our partners should make sure that the data subjects will be able to exercise their rights properly and respond to any relevant requests of the data subjects. These rights include, among others, the following (Articles 15-22 GDPR):

- **Right to information (requirement of transparency).** In accordance with Articles 12, 13 and 14 of the Regulation, the data subject has the right to know the full identity and contact details of those who collect data, either directly from him or indirectly (e.g., through cookies of websites visited). The data subject must also know exactly what data is collected by the data controller, for what purpose, for how long it is kept and to which recipients it is transmitted, if any. The information to the user must be provided in plain language, without any ambiguities, terms, and conditions and without any unclear legal or technical terms.
- **Right to access, erasure, rectify or restrict data.** The data subject has the right (Articles 12 and 15 of the Regulation) to receive full and thorough information as well as oral or written confirmation of the processing, description of the purposes, categories of data and recipients. In addition, the possibility and procedure for submitting a request for rectification, erasure, restriction of processing and for lodging a complaint with the competent authority should be made known to the data subject. The data subject should also be provided with a copy of his/her data. Moreover, the data subject has the right (Articles 5(1)(d), 16 and 17 of the Regulation) to request, within a reasonable period of time, the correction of inaccurate personal data and the completion of incomplete data. The data subject shall bear the burden of proving his or her true identity. The data subject has the right (Articles 5(1)(d), 17 and 19 of the Regulation) to request the erasure of his/her personal data without having to invoke any prejudice to him/her. This is a relative and not an absolute right which might be infringed when there is an overriding legal ground for the retention of personal data. Ultimately, the data subject has the right (Articles 4(3), 12, 18 and 19 of the Regulation) to request the restriction of the processing of his/her data when he/she contests its accuracy, when the processing is unlawful or when the data is no longer needed by the controller.
- **Right to data portability to other data controller** – this is a right to receive the personal data concerning the data subject, which he/she has provided to a controller, in a structured, commonly used, and machine-readable format and have the right to transmit this data to another controller.
- **Right to object against further processing.** In principle, Article 21 GDPR prescribes that the data subject has the right to object the further processing of their data at any time, unless the controller has a justified aim to continue the processing, which overrides the rights and interests of the data subject. It should be up to the data controller to prove that its compelling legitimate interests possibly override the interests or fundamental rights and freedoms of the data subject. However, it is very difficult to prove that processing of data overrides the rights and interests of the data subject. The situations in which such overriding can occur is when there is an epidemic, threat to public health, threat to national security, etc.

3.3 Data protection policy

Under Article 24 of the GDPR, the data controller shall implement data protection policies as part of appropriate technical and organizational measures to demonstrate compliance where proportionate in relation to processing activities. In this sub-section, some proposed measures will be presented, in order for the compliance of the project with the data protection related provisions to be assured.

3.3.1 Data mapping

Data mapping is crucial to the success of many data processes. In the scope of data protection regulations, the purpose of data mapping is firstly to consider whether personal data, as defined in the GDPR, is being processed and if so, to further examine other factors related to data protection principles such as for example the lawfulness of the processing, the respect of the subjects' rights and possible threats to the subject's rights and freedoms related to any act of processing.

In order to achieve this, the TANGO Data Management Plan Questionnaire, Version 1.0, as presented in **Annex 1**, was sent to the partners on 30 November 2022, asking them to provide information about the data expected to be handled during the project, how such data would be handled, and how the FAIR principles would be taken into consideration. In the questionnaire there are also questions related to the handling of research outputs other than data, as well as allocation of responses and ethics matters.

After the evaluation of the answers received, we were able to witness for starters whether any personal data is being/will be processed and if this is the case, whether further measures should be taken regarding this. The answers are also an indicator of the knowledge that TANGO partners have related to data protection principles, the level of their compliance in relation to their project activities and the measures they have implemented (or plan to implement) in order to achieve data security and integrity. Thus, it is important that the partners answer such questions while having the necessary knowledge regarding basic GDPR provisions, as provided in this section, in order to ensure that their answers are informed. As mentioned in **Section 2** above, the answers received in response to the questionnaire sent out in November constitute initial input, and data management processes, including obtaining further information about partners' data-related practices, will iteratively continue during the project.

3.3.2 Data protection policy

As any type of technical and organizational measures to demonstrate compliance, the data protection policy shall be implemented considering the nature, scope, context, and purposes of processing, as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons. In other words, the data protection policy is a set of principles, rules, and guidelines that informs how the organization will ensure ongoing compliance with data protection laws. The policies should recognize the data protection principles and the rights of individuals set out by the GDPR and explain how they are put into practice in relation to the processing carried out by the organization.

The template provided in **Annex 2** is to be implemented and detailed by project partners processing personal data with regards to their activities in TANGO. The partners should make sure that they implement those example provisions and guidelines throughout any act of processing of personal data during the project. They should be properly informed regarding basic GDPR rules and principles, have the necessary knowledge to deal with any data protection related issues they may face and most of all, be willing to seek for guidance or help in any data protection related case.

3.3.3 Data protection officers

Articles 37-39 of the GDPR refer to the appointment of a data protection officer and describe in which circumstances the appointment of such an officer is recommended and necessary. Appointment of a DPO is necessary if data is processed by a public authority, or if the core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale, or if the core activities consist of processing on a large-scale special categories of data.

The partners of the TANGO project are obligated for starters to record an internal analysis in order to determine whether a DPO should be appointed, so that they are able to prove that consideration was given to the nature, scope, context, and purposes of the processing, as well as the risks of different probability of occurrence and severity to the rights and the freedoms of natural persons. This analysis is part of the documentation required under the principle of accountability, could be requested by the supervisory authority, and should be updated when deemed necessary. However, the TANGO partners who are going to process data falling under the special category of personal data (e.g., biometric data) and/or those who carry out regular and systematic monitoring of data subjects on a large scale, are

obliged under the GDPR to appoint a DPO (such cases are for example the piloting activities described in WP7 of the Grant Agreement).

Moreover, the term of "regular and systematic monitoring of data subjects" provided in the DPO provisions may also be a factor on this project. According to the Article 29 Working Party ("WP29") "Guidelines on Data Protection Officers ('DPOs')", regular monitoring is considered to be the process which is ongoing or occurring at particular intervals for a particular period, recurring or repeated at fixed times or constantly or periodically taking place. Systematic monitoring of data subjects could be the process occurring according to a system, pre-arranged, organised, or methodical, taking place as part of a general plan for data collection or carried out as part of a strategy. [22]

To conclude, we suggest the partners to carefully examine whether the appointment of a DPO is required in order to assure that the requirements of Articles 37-39 of the GDPR are being met.

3.4 Data management and measures

With the current sub-section, we will refer to the way personal data could be collected during the course of the project, how this will be managed, described, analysed, and stored and what mechanisms will be used to share and preserve data.

3.4.1 Data processing principles

Data collection, management and in general processing throughout the duration of the TANGO project, will be guided through the following principles/measures:

- Data could be collected/processed mostly in an **anonymized form**. In this case, the questionnaires, interview guidelines and other used instruments, where possible, must not contain questions, whose answers could lead to the participant's identity – alone or in combination with other answers.
- For the purposes of individual tasks, data could be pseudonymized instead. In this case, the partners should justify their action and data subjects should be informed and provide the partners with the necessary consent, according to Article 6 GDPR, to the extent that this is the legal basis for processing. Data should be anonymized by the time data processing is finished.
- The legal basis of processing of project partners' personal data (names, communication information, etc.) for the purposes of the project (for example in order to communicate and cooperate with other partners of the project) could be performance of the contract, or another legal basis as appropriate.
- The legal basis of processing of personal data of any of the participants to the project (stakeholders, citizens participating to pilot activities, etc.) could be consent according to Article 6 GDPR.
- Partners are strongly advised to refrain from data processing of special categories of personal data of Article 9 or 10 GDPR. In case the processing of such personal data is necessary for the purposes of the project, the partners should be provided with the necessary consent, in which the data subjects should be also informed regarding the necessity of such acts of processing.
- The **anonymity**, where possible, and the **privacy of participants** (stakeholders, citizens participating to pilot activities, etc.) must be respected. Personal information must be kept confidential and transferred to other project partners only if this necessary for a specific task related to the project. In this case, data subjects should be informed regarding the possibility of such acts of processing while providing the partners with their consent. Guarantees of confidentiality and anonymity given to the participants must be honoured, unless there are clear and overriding reasons to do otherwise.
- In case that the participants must be registered at the TANGO platform, they must not be registered with their name, if possible. E.g. an ID-code could be applicable instead of it. That **takes into account privacy considerations** and further, the ID-code helps to match answers of questionnaires and the data collected at the platform by the user.
- The data subjects themselves have their **data sovereignty**. Although the data has been provided by them and is being processed for the purposes of the project, the data subjects retain control over their data. So, for example, in case the data subjects request the deletion of their data, this has to be done without any undue delay.

- The participant is allowed to change/limit the access authorization of their data collected at the TANGO platform.
- All researchers have the duty of maintaining confidentiality of the collected data.
- The integrity of stored, processed and published data must be ensured by the researchers and the project consortium.
- Data that is collected by the participant at the TANGO platform must be treated with care:
  - Participants must be informed that the data could be used for the project.
  - Participants must be informed in which way the data could be used.
  - The participants must be informed when the collected data will be deleted.
  - Appropriate measures, namely cryptography and physical security measures, must be taken by the partners to process data in secure manner.
  - In case the participant withdraws from the pilot, the collected data at the platform must be deleted or the access to them must be impossible for others, without any undue delay.

3.4.2 Security of processing

The GDPR does not stipulate exact ways of data security, rather it gives minimum recommendations and urges the data controllers to make decisions on which technical and organizational measures to take depending on the state of art, the cost to implement, the varying likelihood of risk, but also the fundamental rights and freedoms at stake. Technical and organisational measures are for instance (Article 32 GDPR):

- The pseudonymization and encryption of personal data. Pseudonymization of data means replacing any information which could be used to identify an individual with a pseudonym, or, in other words, a value which does not allow the individual to be directly identified. It is a technical and organizational measure to ensure non-attribution to an identified or identifiable person and, according to the GDPR, **pseudonymized data are "personal data" since they can indirectly help identify the data subject with the use of additional information.** On the contrary, anonymized data (anonymization is the process of removing personal identifiers, both direct and indirect, that may lead to an individual being identified) cannot help identifying the data subject since all the personal identifiers have been removed and as such, is not considered to be "personal data" according to the Regulation.
- The ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident.
- A process for regularly testing, assessing, and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

The partners, in order to be able to demonstrate compliance with this Regulation, should establish internal policies and implement measures, which specifically respond to the principles of data protection by design and by default. Data protection by design means that implementation of appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data protection principles, such as data minimization, shall be done in an effective manner. The necessary safeguards into the processing shall be integrated from the very outset to meet the requirements and protect the rights of data subjects (Article 25 GDPR). Data privacy by default means that only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of its storage and its accessibility.

TANGO could consider the implementation of the following techniques [23]:

- **Directory replacement**: A directory replacement method involves modifying the name of individuals integrated within the data, while maintaining consistency between values, such as "postcode + city".
- **Scrambling**: Scrambling techniques involve a mixing or obfuscation of letters. The process can sometimes be reversible. For example: "Annecy" could become Yneanc.

- **Masking**: A masking technique allows a part of the data to be hidden with random characters or other data. For example, pseudonymization with masking of identities or important identifiers. The advantage of masking is the ability to identify data without manipulating actual identities.
- **Personalized anonymization**: This method allows the user to utilize their own anonymization technique. Custom anonymization can be carried out using scripts or an application.
- **Blurring**: Data blurring uses an approximation of data values to render their meaning obsolete and/or render the identification of individuals impossible.

While the project evolves, the consortium will compare the proposed techniques and decide to implement one or more techniques according to the needs of the project.

3.4.3 Data minimization

The data minimization principle comprises that data has to be adequate, relevant and limited to what is necessary for the purposes for which it is processed. This implies that:

- Data collected, processed, analyzed, and archived should not be held or further processed, unless this is essential for reasons that were stated in advance.
- Data collection and processing should only include as much data as is required to successfully answer the research question(s).
- Data collected for one purpose can be repurposed only under strict restrictions. The processing of personal data for purposes other than those for which it was initially collected should be allowed only where the processing is compatible with the purposes for which the personal data was initially collected. In such a case, no legal basis separate from that which allowed the collection of the personal data is required. Moreover, further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be considered to be compatible lawful processing operations in such cases. In order to ascertain whether a purpose of further processing is compatible with the purpose for which the personal data was initially collected, the controller, after having met all the requirements for the lawfulness of the original processing, should take into account, inter alia: any link between those purposes and the purposes of the intended further processing; the context in which the personal data has been collected, in particular the reasonable expectations of data subjects based on their relationship with the controller as to their further use; the nature of the personal data; the consequences of the intended further processing for data subjects; and the existence of appropriate safeguards in both the original and intended further processing operations (Recital 50 GDPR).

In the TANGO project, the following guidelines are provided to the partners in order for them to conclude whether the data minimization principle is respected:

- When collecting personal data, ask yourself for which purpose you collect the data, how you are planning to use the data, and whether there is a way of achieving this purpose without having to collect the personal data. Document the choices you make in this process.
- Only collect the personal data that is strictly necessary to achieve the purpose, i.e., answering the research question(s). The partners should not collect personal data that is not compatible to the purposes of the processing.
- It is advisable not to keep the personal data stored longer than necessary to achieve the purpose, i.e., answering the research question(s), being able to prove validity of research outcomes, complying with legal obligations, etc.
- De-identification of personal data reduces the chance of identification. Anonymization (see previous sub-section) is the process in which you delete all information that may lead to identification of an individual. Consider indirect indicators and combinations of indicators as well, as these may lead to identification as well. Once personal data is properly anonymized, the data does not fall within the scope of the GDPR anymore.
- Another form of de-identification is pseudonymization (see previous sub-section), which offers a (temporary) solution when personal data is necessary to keep (for instance for longitudinal research or accounting for scientific integrity), but the personal data itself is redundant in the daily routine of

processing and analyzing data. Pseudonymization refers to the process of replacing personal identifiers with codes that are stored in a different file on a different location. Keep in mind that pseudonymized data still remains personal data and therefore the GDPR still applies to this data.

- Repurposing personal data becomes an issue in the case the purpose is formulated in a (too) restrictive way in the informed consent procedure (e.g., "for this research project" or "by the involved researchers"). If you reasonably expect repurposing of personal data in the near future, we advise you to make sure your consent is formulated wide enough (e.g., "for research purposes" or "by researchers employed by a scientific organization"). However, data may only be processed for specified and explicit purposes (Article 5 GDPR).

3.4.4 Data breaches notification obligation

According to Article 33 GDPR, in the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay. The processor shall notify the controller without undue delay after becoming aware of a personal data breach. The notification referred above shall at least:

- Describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned, and the categories and approximate number of personal data records concerned.
- Communicate the name and contact details of the DPO or other contact point where more information can be obtained.
- Describe the likely consequences of the personal data breach.
- Describe the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects.

3.5 Data protection impact assessment

One of main elements of the GDPR, introduced in Article 35, is the need to perform a DPIA in specific situations. Although not specifically described in the GDPR, a DPIA is considered as a process designed to describe the data processing and assess its necessity and proportionality. DPIAs are designed to help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data by assessing them and determining the measures to address them. DPIAs are important tools for accountability, as they help controllers not only to comply with requirements of the GDPR, but also to demonstrate that appropriate measures have been taken to ensure compliance with the Regulation.[5] In other words, a DPIA is a process for building and demonstrating compliance and to avoid possible consequences of non-compliance which can cause a fine up to 4% of worldwide turnover for a company.

The GDPR places obligations on both the data controller, which "alone, or jointly with others, determines the purposes and means of the processing of personal data", and the data processor, who processes personal data on behalf of the controller. However, the subject responsible for carrying out a DPIA is data controller. [24] If the processing is wholly or partly performed by a data processor, the processor should assist the controller in carrying out the DPIA and provide any necessary information.

When a DPIA is required

A DPIA [25] is required under the GDPR any time a new project is beginning that is likely to involve "a high risk" to other people's personal information. The variety of discussion running nowadays related to impact assessments and the related GDPR direction around DPIAs demonstrate the difficulties to

[5] See also Recital 84 GDPR: "The outcome of the assessment should be taken into account when determining the appropriate measures to be taken in order to demonstrate that the processing of personal data complies with this Regulation".

define borders which separate the need to do mandatory assessment (DPIA) from the simple adherence to GDPR principles. In general, the suggestion made by authorities is to run the assessment if not sure to be directly or indirectly involved in one of the cases which require that.

Furthermore, it is important to maintain compliance during processes, so an assessment will be a periodic action to be performed. The aim of those periodic assessment documents is to provide evidence of methodology, procedures and related outcomes that will be part of the compliance process activated by TANGO partners to provide their specific assessments related to Legal, Social, Ethics and Liability issues.

In line with the risk-based approach embodied by the GDPR, carrying out a DPIA is not mandatory for every processing operation. [24] The GDPR demands that a DPIA be carried out "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons."

However, there is no "silver bullet" method for carrying out impact assessments [26]: "What matters is the choice of an appropriate assessment method allowing for the best understanding and treatment of possible consequences of the envisaged initiative. These methods can range from qualitative or quantitative risk management to scenario planning, to scientific foresight, supported by a compliance check with relevant legal and otherwise regulatory requirements (e.g., technical standards)." [26] The GDPR sets out the minimum features of a DPIA (Articles 35(7) and Recitals 84 and 90):

- A description of the envisaged processing operations and the purposes of the processing;
- An assessment of the necessity and proportionality of the processing;
- An assessment of the risks to the rights and freedoms of data subjects;
- The measures envisaged to "address the risks" and "demonstrate compliance with this Regulation". [24]

A DPIA may concern a single data processing operation or could be used to assess multiple processing operations that are similar in terms of nature, scope, context, purpose, and risks. [24] Depending on the specific implementation of TANGO project, it should be decided what processing operation falling into the DPIA requirement can be deemed similar. Moreover, a DPIA can also be useful for assessing the data protection impact of a technology product, for example a piece of hardware or software, where this is likely to be used by different data controllers to carry out different processing operations. [24]

As indicated above, according to the Regulation, "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data". While this passage makes it clear that a DPIA is required by law under certain conditions, it is unhelpfully light on specifics. To help clarify the situation, here are some concrete examples of the types of conditions that would require a DPIA: [25]

- If you are using new technologies;
- If you are tracking people's location or behaviour;
- If you are systematically monitoring a publicly accessible place on a large scale;
- If you are processing personal data related to "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation";
- If your data processing is used to make automated decisions about people that could have legal (or similarly significant) effects;
- If you are processing children's data;
- If the data you are processing could result in physical harm to the data subjects if it is leaked.

Additionally, the WP29 "Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679" provide the following criteria that shall be considered to define the need for DPIA:

- Evaluation or scoring;
- Automated decision-making with legal or similar significant effect;
- Systematic monitoring;
- Data processed on a large scale;
- Sensitive data or data of a highly personal nature;
- Matching or combining datasets;
- Data concerning vulnerable data subjects;
- Innovative use or applying new technological or organizational solutions;
- When the processing in itself "prevents data subjects from exercising a right or using a service or a contract".

In other cases, where the high-risk standard is not met, it may still be prudent to conduct a DPIA to minimize your liability and ensure best practices for data security and privacy are being followed during the progress of the project.

Why DPIA in TANGO

Although the project may not be running "systematic" actions as described in GDPR, being a research project in which some actions carried out (e.g., piloting activities, dissemination, etc.) could go under some of criteria that flag the need of DPIA. In view of these, we suggest all partners involved as data controllers to complete a dynamic DPIA.

The GDPR makes it clear (Article 35 and Recitals 89 [27] and 91[28]) that the use of a new technology, defined in "accordance with the achieved state of technological knowledge" (Recital 91), can trigger the need to carry out a DPIA. This is because the use of such technology can involve novel forms of data collection and usage, possibly with a high risk to individuals' rights and freedoms. Indeed, the personal and social consequences of the deployment of a new technology may be unknown. A DPIA will help the data controller understand and treat such risks.

In the early months of the project this is considered to be preliminary, as there is not yet a detailed view of data use along the project. A DPIA has to take place before any acts of the data processing start but can always be updated if the circumstances, under which it had taken place, change during the data processing. Since a DPIA is part of T2.6 and will be carried out during the following stages of the project and before the actual piloting activities take place, the partners should be willing to cooperate with the partner in charge of this task in order to help predict and preliminarily prevent any possible issues that may arise in the future and accomplish the goal of data security and integrity.

3.6 Ethical issues and societal concerns in TANGO

For all activities funded by the EU, ethics is an integral part of research from beginning to end, and ethical compliance is seen as pivotal to achieve real research excellence. There is clear need to make a thorough ethical evaluation from the conceptual stage of the proposal not only to respect the legal framework but also to enhance the quality of the research. Ethical research conduct implies the application of fundamental ethical principles and legislation to scientific research in all possible domains of research. The process to assess and address the ethical dimension of activities is called the Ethics Appraisal Procedure, in order to ensure that the provisions on ethics are respected.

3.6.1 General provisions

The partners must carry out the action in compliance with the ethical principles (including the highest standards of research integrity) and the applicable international, EU and national law. They must respect the fundamental principle of research integrity, as set out in the European Code of Conduct for Research Integrity. While carrying out with the project, the partners should:

- Respect human dignity and integrity;
- Ensure honesty and transparency towards research subjects and notably get free and informed consent (as well as assent whenever relevant);

- Protect vulnerable persons;
- Ensure privacy and confidentiality;
- Promote justice and inclusiveness;
- Minimise harm and maximising benefit;
- Share the benefits with disadvantaged populations, especially if the research is being carried out in developing countries;
- Maximise animal welfare, in particular by ensuring replacement, reduction and refinement in animal research;
- Respect and protect the environment and future generations.

According to Article 19(1) of the Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, "particular attention shall be paid to the principle of proportionality, to the right to privacy, the right to the protection of personal data, the right to the physical and mental integrity of a person, the right to non-discrimination and to the need to ensure protection of the environment and high levels of human health protection".

The partners should be capable to predict the cases in which ethical issues may be raised. Before the beginning of an activity as such, each partner must have obtained any ethics committee opinion required under national law and any notification or authorisation for activities raising ethical issues required under national and/or European law.

If a partner breaches any of the obligations above, the penalties foreseen in Chapter 5 of the "GRANT AGREEMENT Project 101070052 — TANGO" will be in place.

3.6.2 Research integrity

In order to ensure the necessary level of research integrity, the partners must follow principles listed below and ensure that the people carrying out research tasks comply with the European Code of Conduct for Research Integrity. The fundamental research integrity principles are:

- Reliability in ensuring the quality of research reflected in the design, the methodology, the analysis and the use of resources;
- Honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair and unbiased way;
- Respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment;
- Accountability for the research from the idea to publication, for its management and organisation, for training, supervision, and mentoring, and for its wider impacts.

3.6.3 Ethics and data protection

In this sub-section, we will mostly refer to ethical issues that may be raised and are related to data protection. Other research sectors in which ethical issues may arise but we do not consider them related to the current project are research on human embryos and foetuses, human cells, or tissues, human science research and research related to the environment, the health and safety of people or animals.

Data protection is both a central issue for research ethics in Europe and a fundamental human right. The right to data protection is enshrined in the EU Charter of Fundamental Rights and the Treaty on the Functioning of the European Union, which give effect to individuals' right to privacy by providing them with control over the way information about them is collected and used.

In research settings, data protection imposes obligations on researchers to provide research-data subjects with detailed information about what will happen to the personal data that they collect. It also requires the organizations processing the data to ensure the data is properly protected, minimized, and destroyed when no longer needed. Depending on the setting or information in question, the failure to protect personal data against loss or misuse can have devastating consequences for the data subjects. It may also have serious legal, reputational and financial consequences for the data controller and/or processor.

Many recent examples of unethical research practices have involved the unauthorized collection and/or (mis)use of personal data, resulting in enforcement action by regulators.

It should be highlighted that the fact that research is legally permissible according to data protection legislation does not necessarily mean that it will be deemed ethical. This especially applies in cases in which the processing of personal data raises higher ethics risks, when it involves: processing of "special categories" of personal data (formerly known as "sensitive data"), processing of personal data concerning children, vulnerable people or people who have not given their consent to participate in the research; complex processing operations and/or the processing of personal data on a large scale and/or systematic monitoring of a publicly accessible area on a large scale; data processing techniques that are invasive and deemed to pose a risk to the rights and freedoms of research participants, or techniques that are vulnerable to misuse; and collecting data outside the EU or transferring personal data collected in the EU to entities in non-EU countries.

In case of higher-risk data processing, a detailed analysis of the ethics issues raised by the project methodology is needed, which should comprise an overview of all planned data collection and processing operations, identification, and analysis of the ethics issues that these raise and an explanation of how you will mitigate these issues in practice.

Processing of "special categories" of personal data, according to Article 9 GDPR, should be treated with caution since, apart from the strict legislative provisions of GDPR, ethical issues may arise too. Project partners who are due to process special categories of data should first of all be able to justify and determine the necessity of their action according to the principles of accountability and purpose limitation, taking into consideration also the nature of the project and the individual tasks.

The process of securing the explicit and informed consent of data subjects is of utmost importance. The partners of the project should explain to research participants-data subjects what the research is about, what their participation in the project will entail and any risks that may be involved. Only after they have conveyed this information to the participants – and they have fully understood it – we can seek and obtain their express permission to include them in the project.

The consent should be provided according to the GDPR. This requires consent to be given by a clear affirmative act establishing a freely given, specific, informed, and unambiguous indication of the subject's agreement to the processing of their personal data. This may take the form of a written statement, which may be collected by electronic means, or an oral statement. Data subjects should be properly informed with the use of clear and plain language regarding their rights, all the data processing activities, the principles of data processing and the way they can withdraw their provided consent at any time. As a minimum, this should include the identity of the data controller and, where applicable, the contact details of the DPO, the specific purpose(s) of the processing for which the personal data will be used, the subject's rights as guaranteed by the GDPR and the EU Charter of Fundamental Rights, in particular the right to withdraw consent or access their data, the procedures to follow should they wish to do so, and the right to lodge a complaint with a supervisory authority, information as to whether data will be shared with or transferred to third parties and for what purposes and how long the data will be retained before it is destroyed.

Moreover, whenever the legal basis of the processing is the consent of the data subject, our partners should limit data processing to the purposes under which the consent was provided by the data subject. Repurposing or re-use of previously obtained personal data is for starters prohibited since it is considered to be a clear violation of Article 5(1)(a) GDPR, although, as mentioned above, the processing of personal data for purposes other than those for which it was initially collected should be allowed only where the processing is compatible with the purposes for which the personal data was initially collected. In such a case, no legal basis separate from that which allowed the collection of the personal data is required.

If our partners are going to use data that is publicly available, they must provide details of the source(s) and confirm that the data is openly and publicly accessible and may be used for research purposes. In the case data from social media networks will be used, our partners must assess whether those people actually intended to make their information public. Overall, in such cases, it should be examined whether the data subject had any reasonable expectation of privacy while making their data publicly available,

alongside other factors, such as the terms of use and the privacy policy of the data controller of such platform.

Finally, ensuring data safety and integrity is a mandatory in terms of ethics. We have already highlighted to our partners the need of implementing the necessary technical and organizational measures in order to ensure data security and integrity. The use of anonymization and pseudonymization techniques during the collection and in general processing of personal data is of utmost importance for starters, while at the same time the necessary measures should be taken also to ensure that data are securely stored. Considering the "risk-based approach" of the GDPR, the necessary security measures should be implemented depending on the degree of the risk to the rights of the data subjects, and thus our partners should be accordingly informed and able to encounter any possible threats that may arise.

To conclude, our project partners should proceed with caution regarding any data processing actions that may raise ethical issues, such as the processing of special categories of data or the repurposing of data. Other potential cases in which ethical issues may arise could be the processing of data of children or vulnerable people but, taking into consideration the nature of each task of the project, our partners should refrain from any act of processing of such personal data. The need of completion of a dynamic DPIA by TANGO project partners, as described in T2.6 in detail, could help avoid preliminarily any ethical issues may rise in the future.

4 Responsible Research and Innovation Practices

Promoting RRI is one of the operational objectives of HE. [29] This term broadly entails that the public (i.e., all relevant actors/stakeholders) is engaged in the research process to better align the goals and outcomes of research with societal needs and to address societal challenges. RRI practices, reflected in the HE Regulation, the HE Program Guide, and the HE Strategic Plan (2021-2024), [30] relate to, for example, open science (including open access), public engagement, ethics and integrity, gender equality and inclusiveness, and societal impact.

This section primarily aims to make a preliminary identification of aspects related to RRI that shall be considered by the TANGO consortium and, as appropriate, guide partners' individual and joint efforts during the project. It also provides, to the extent possible at this stage, an initial, high-level overview of steps anticipated to be taken by TANGO partners during their project activities in line with these aspects. Further information will be provided as part of the work to be carried out in other tasks/WPs, as indicated in the sub-sections below, and will be reported in the respective deliverables. An update, including more concrete discussion on the implementation of RRI practices by TANGO partners, will be presented in the final version of the DMP and research ethics deliverable (deliverable D1.3), in month 32 of the project.

Overall, TANGO aims to deliver positive societal impact. The project will consider and address societal needs, notably that of data availability and sharing, in light of recent societal developments and in line with the European Strategy for data, [31] in a manner that respects the fundamental rights of citizens (particularly the right to privacy and data protection) and that places the user at the centre of the decision-making regarding such data sharing. The TANGO environment will be designed to increase trust, ensure compliance and privacy, and allow "green" data operations to the benefit of not only organizations but, importantly, also of citizens interacting with them in their daily lives. In particular, TANGO solutions will benefit citizens through the offering of strong privacy-preserving identity management features which will provide user-centric data control along with a secure and robust mechanism for data sharing. Citizens, and ultimately public and private organizations, will be encouraged to actively engage in data sharing, in a manner that is user-centric, privacy-preserving and trustworthy.

A multidisciplinary approach has been opted for in the project, with partner organizations comprising a balanced variety of societal actors – universities and research centres, think tanks, business and technology associations and businesses in various domains. All these actors will work together during the project's research and implementation phases to make sure that the proposed solutions are in line with the values, needs and expectations of society. In addition, an analysis and assessment of citizens' and organizations' acceptance of the TANGO solution in six use cases with a high societal, environmental, and financial impact on citizens and the data economy – namely, smart hospitality, autonomous vehicles, smart manufacturing, financial institutes, public administration, and retail – will be undertaken. Finally, social issues of the TANGO framework are anticipated to be assessed as part of WP2/T2.6 carrying out a comprehensive Privacy, Ethical, Social and Legal Impact Assessment.

The sub-sections below provide information related to specific RRI areas, notably open science (**Sub-section 4.1**), public engagement (**Sub-section 4.2**), ethics and integrity (**Sub-section 4.3**) and gender equality and inclusiveness (**Sub-section 4.4**).

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. [1][21] It encompasses – but goes beyond – open access, which is access – free of charge for the end user – to research outputs resulting from the project. [21] 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. [29] 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, more easily 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 validate/re-use research data; and (iv) digital or physical access to results needed to validate the conclusions of scientific publications, unless exceptions apply. [21][1]

TANGO 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 include the following: *First*, the first version of the TANGO DMP has been set out in this deliverable, and the DMP will be updated in the course of the project, with a final version being included in D1.3. *Second*, the FAIR principles have been presented in **Section 2** of this document and will be considered by TANGO partners to decide on the appropriate measures to be taken with regards to making data and other research outputs FAIR. *Third*, open access to peer-reviewed scientific publications will be pursued. Appropriate arrangements and decisions, including with regards to the trusted repository to be chosen (e.g., Zenodo) and the necessary open licenses that will accompany each publication (e.g., Creative Commons), as well as the metadata of deposited publications, will be considered by partners, taking into account the obligations undertaken under the GA. It is foreseen that publications, journal articles, conference papers and presentations based on project results will generally be open access, and public deliverables produced in the course of the project will be made available through the TANGO website.

Fourth, open access to research data shall be provided in line with the premise "as open as possible, as closed as necessary". As appropriate with regards to each dataset, and to the extent that this would not be against the partner's legitimate interests (including commercial exploitation) or any other constraints (such as data protection rules, privacy, confidentiality, trade secrets, security rules or IPRs), TANGO partners shall ensure that data is deposited in a trusted repository, accompanied by the necessary licenses, as soon as possible after data production and at the latest by the end of the project. As appropriate, TANGO partners shall also ensure that data underpinning a scientific publication be deposited at the latest at the time of publication, in line with standard community practices and accompanied by the appropriate license. Information about any research output or any other tools and instruments needed to re-use or validate the data, as well as open metadata of deposited data shall also be provided. TANGO partners, in alignment with the European Open Science Cloud ("EOSC"), shall seek to provide seamless access and reliable re-use of research data to researchers, innovators, companies and citizens through a trusted and open distributed data environment.

In addition to the above mandatory open science practices, recommended open science practices shall also be considered and may be adopted by partners, jointly and/or individually, as appropriate. Such practices may include but are not limited to involving relevant knowledge actors (including citizens, civil society, and end-users) early, open sharing of research, and management of research outputs beyond research data (e.g., software tools, models, etc.). [1]

By way of example, relevant knowledge actors, such as stakeholders and citizens, are expected to be involved in project activities, as explained in **Sub-section 4.2** below. Research activities undertaken as part of the TANGO project will be shared with third parties through publications and other dissemination activities further explained in D8.1. Research outputs other than data are also expected to be managed in accordance with the FAIR principles, as outlined in **Sub-section 2.4** above.

As the project is currently at an early stage, open access and open science practices will be further defined and specified at a later stage, taking into account the work carried out under various WPs/tasks.

For example, IPR and innovation management considerations explored under WP8/T8.3, the partners' exploitation plans prepared as part of WP8/T8.2, as well as the commercial roadmap for the TANGO solutions set out under the remit of WP8/T8.4, are expected to have an impact on the concrete steps to be taken to implement open science practices under TANGO. A more detailed update as to the open science and open access practices adopted by TANGO partners will be provided in D1.3.

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 taken into account, that science education is promoted, and that scientific knowledge is publicly accessible. [21][29]

The TANGO consortium, acknowledging the significance of public engagement, envisages taking steps in line with this RRI practice, by, e.g., communicating research results to the public and by involving stakeholders in its research activities. For example, citizens are anticipated to be involved in certain demonstrations to be executed under WP7 (the scope of citizen involvement will be determined as the project progresses), providing input that can be used to inform the validation and evaluation of the TANGO solution, demonstrate its value proposition, understand user satisfaction and experience with the framework and identify meaningful improvements. External stakeholders (e.g., technical experts, lawyers, privacy advocates, human rights experts and end users) will also be involved in the Privacy, Ethical, Social and Legal Impact Assessment to be undertaken under WP2/T2.6, by participating in the interactive workshops to be organized and interviews to be carried out. T2.6 will also assess social issues of the TANGO framework.

Moreover, TANGO's activities will be disseminated and communicated to the public through a dedicated website, social media channels, blogposts, and a newsletter. Furthermore, participation in and/or organization of events will allow TANGO partners to disseminate their findings and communicate their project-related activities to the research community, the industrial world and the public. Scientific publications constitute an additional channel that will be leveraged by TANGO partners. Further information on dissemination and activities conducted with regards to communication channels used (including dissemination) can be found in D8.1, which, building on the activities of WP8/T8.1, lays out the first version of the TANGO Dissemination and Communication Plan, acting as guide for dissemination and communication activities during the project.

With the goal of promoting public engagement in mind, TANGO partners will further develop and refine, as the project progresses, how this goal is to be achieved to the best of their abilities. Further information about the RRI practices related to public engagement will be provided in D1.3.

4.3 Ethics and integrity

For EU-funded activities, the ethical dimension is an integral part of the research lifecycle, and ethical compliance is of utmost importance for achieving research excellence. [1] The TANGO consortium recognizes the importance of ethics and integrity in research, and the necessity of ethical compliance, and partners shall carry out their actions in accordance with the highest ethical standards and legal requirements. To this end, as explained in **Sub-section 2.6**, a dedicated ethics partner has been appointed by the consortium, and the first version of the research ethics and compliance protocol has been defined and set out in **Section 3** above. It will be further developed, iteratively adjusted, and monitored throughout the project. Ethical requirements, including but not limited to those relating to data protection and privacy, will be taken into account during the project. In addition, a comprehensive impact assessment covering ethical, as well as privacy, social and legal issues will be carried out as part of WP2/T2.6. This impact assessment will identify, among others, the ethics-related themes of the TANGO environment, identify any risks and suggest possible solutions and mitigation measures, as necessary

and appropriate. Finally, certain ethics issues will be addressed under WP9, which sets out and will ensure compliance with specific ethics requirements.

In addition, the fundamental principles of research integrity as set out in the European Code of Conduct for Research Integrity shall be respected by TANGO partners. [32] Consequently, TANGO partners shall ensure compliance with the principles of: reliability in ensuring the quality of research reflected in the design, the methodology, the analysis and the use of resources; honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair and unbiased way; respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment; and accountability for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider impacts. Furthermore, TANGO partners shall make sure that persons carrying out research activities abide by the good research practices enshrined in the Code.

4.4 Gender equality and inclusiveness

This cross-cutting principle embedded in HE encompasses the existence of a gender equality plan within public bodies, research organizations and higher education establishments, 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"). [21][1][29]

The TANGO consortium shall duly consider how to integrate relevant considerations into its work, in line with this RRI practice. Partners shall consider, for instance, having a balanced representation of women and men in co-design activities, such as workshops, and having regard to gender issues where relevant and appropriate (e.g., when it comes to the training and evaluation process of AI/ML algorithms, as per existing guidance). Furthermore, research activities under the project will overall be carried out by a gender-balanced team of researchers. Information regarding this RRI practice will be provided in the final version of the DMP set out in D1.3.

5 Conclusions

This document constitutes the first version of the Data Management Plan and Research Ethics deliverable of the TANGO project. To this end, it has provided, among others, an initial overview of how data and other research outputs are expected to be handled by partners, as well as the ethics guidelines, including but not limited to data protection guidelines and considerations, that shall guide partners' activities. In addition, this document identified relevant RRI practices, such as open science and public engagement, to be considered during the project. Overall, this document relates to all WPs/tasks of the project, and TANGO partners shall take its content into account when carrying out their research activities.

As explained above, the project is currently at an early stage, meaning that practices, processes, and details need to be further elaborated upon as the project progresses. Certain matters related to data management, ethics and/or RRI practices, will be addressed as part of other tasks and deliverables of the TANGO project. An updated version of the DMP, the research ethics protocol and RRI practices will be provided in month 32 of the TANGO project, in D1.3.

6 References

[1] 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.

[2] 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.

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

[4] Consortium of European Social Sciences Data Archives (CESSDA) (2019), *Adapt your Data Management Plan: A list of Data Management Questions based on the Expert Tour Guide on Data Management*, https://static-archive.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf, retrieved on 03-01-2023.

[5] 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 (General Data Protection Regulation), *OJ L 119*, 4 May 2016, pages 1-88.

[6] Jones, T. (2022), *What Is CIA Triad? – The Backbone Of Information Security*, https://medium.com/nerd-for-tech/what-is-cia-triad-the-backbone-of-information-security-712659c7206e, retrieved on 08-03-2023.

[7] GitHub, *Confidentiality, Integrity, Availability (CIA)*, https://github.com/0xsanny/guides/blob/master/src/pages/security/confidentiality-integrity-availability/index.md, retrieved on 08-03-2023.

[8] Wilkinson, M. D. et al. (2016), *Comment: The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific Data, 3*, 160018.

[9] GO FAIR, *F1: (Meta) data are assigned globally unique and persistent identifiers*, F1: (Meta) data are assigned globally unique and persistent identifiers, https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/, retrieved on 03-01-2023.

[10] GO FAIR, *F2: Data are described with rich metadata*, https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/, retrieved on 03-01-2023.

[11] GO FAIR, *F4: (Meta)data are registered or indexed in a searchable resource*, https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/, retrieved on 03-01-2023.

[12] GO FAIR, *A1: (Meta)data are retrievable by their identifier using a standardised communication protocol*, https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/, retrieved on 03-01-2023.

[13] GO FAIR, *A1.1: The protocol is open, free and universally implementable*, https://www.go-fair.org/fair-principles/a1-1-protocol-open-free-universally-implementable/, retrieved on 03-01-2023.

[14] GO FAIR, *A1.2: The protocol allows for an authentication and authorisation procedure where necessary*, https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/, retrieved on 03-01-2023.

[15] GO FAIR, *FAIR Principles*, https://www.go-fair.org/fair-principles/, retrieved on 03-01-2023.

[16] GO FAIR, *I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*, https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/, retrieved on 03-01-2023.

[17] GO FAIR, *I2: (Meta)data use vocabularies that follow the FAIR principles*, https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/, retrieved on 03-01-2023.

[18] GO FAIR, *I3: (Meta)data include qualified references to other (meta)data*, https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/, retrieved on 03-01-2023.

[19] GO FAIR, *R1: (Meta)data are richly described with a plurality of accurate and relevant attributes*, https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/, retrieved on 03-01-2023.

[20] TANGO, *D1.1 Project Management HandBook*, Zaldívar Iván (2022).

[21] 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.

[22] Article 29 Data Protection Working Party (2017), *Guidelines on Data Protection Officers ('DPOs'), Adopted on 13 December 2016, As last Revised and Adopted on 5 April 2017*, 16/EN WP 243 rev.01.

[23] European Union Agency for Cybersecurity, *ENISA proposes Best Practices and Techniques for Pseudonymisation*, https://www.enisa.europa.eu/news/enisa-news/enisa-proposes-best-practices-and-techniques-for-pseudonymisation, retrieved on 03-03-2023.

[24] Article 29 Data Protection Working Party (2017), *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Adopted on 4 April 2017, as last Revised and Adopted on 4 October 2017, 17/EN WP 248 rev.01.

[25] Wolford, B., *Data Protection Impact Assessment (DPIA)*, https://gdpr.eu/data-protection-impact-assessment-template/, retrieved on 03-03-2023.

[26] Kloza, D. et al. (2017), *Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals, d.pia.lab Policy Brief 2017/1*.

[27] Intersoft consulting, *Recital 89 Elimination of the General Reporting Requirement*, https://gdpr-info.eu/recitals/no-89/, retrieved on 03-03-2023.

[28] Intersoft consulting, *Recital 91 Necessity of a Data Protection Impact Assessment*, https://gdpr-info.eu/recitals/no-91/, retrieved on 03-03-2023.

[29] Council Decision (EU) 2021/764 of 10 May 2021 establishing the Specific Programme implementing Horizon Europe – the Framework Programme for Research and Innovation, and repealing Decision 2013/743/EU, *OJ L 167I* , 12.5.2021, pages 1–80.

[30] European Commission, Directorate-General for Research and Innovation (2021), *Horizon Europe Strategic Plan 2021-2024*, Publications Office, https://data.europa.eu/doi/10.2777/083753, retrieved on 03-01-2023.

[31] European Commission, Shaping Europe's digital future: *A European Strategy for data*, https://digital-strategy.ec.europa.eu/en/policies/strategy-data, retrieved on 03-01-2023.

[32] ALLEA – All European Academies (2017), *The European Code of Conduct for Research Integrity, Revised Edition*, ISBN 978-3-00-055767-5.

Annex 1

In preparation of the TANGO DMP, the TANGO Data Management Plan Questionnaire, Version 1.0 was sent to partners on 30 November 2022, asking them to provide information about the data to be handled during the project, how such data would be handled, and how the FAIR principles would be taken into consideration. The questionnaire also asked questions about the handling of research outputs other than data, as well as allocation of responses and ethics matters. The partners' responses to this questionnaire informed the construction of the first version of the DMP of the TANGO project. This Annex contains the questions included in the TANGO Data Management Plan Questionnaire, Version 1.0.

****

Please indicate the name of the partner providing the responses.

Please identify (by name and persistent identifier) your dataset(s), indicating, to the extent relevant, which dataset will be used for which WP.

1. Data summary

Please describe the data you will collect/generate/use. In doing so, please refer to the **type** of data (e.g., numeric, textual, "electronic document", etc.), the **format** (the way in which the data is encoded for storage; you may also explain whether the format will change from the original to the processed/final data), the expected **size** (e.g. the storage space required and/or number of objects, files, etc.), etc.
Please indicate 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. In you answer, 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). Please also indicate whether data concerning **children or vulnerable adults** will be processed.
Are personal data in the dataset pseudonymised? If not, please explain why this is the case.
Are data in your dataset anonymised? If not, please explain why you need to use personal data (in other words, why you cannot reach the research objectives by using anonymised data).
Please identify the **origin/provenance** of the data (**how, where, by whom and when** the data was/will be collected, produced or obtained). Please include in your answer the methods/modes of data collection (e.g., the instruments, hardware, and software used if new data is collected or produced) and explain how data provenance will be documented.

Will you **re-use existing data** (including, but not limited to, publicly available data)? If so, specify the source of the data, providing information about the owner/provider of the dataset. Please also explain how you will re-use them and if there are any constraints on the re-use of such data.
If you have considered but discarded the re-use of existing data, please explain the reasons for that.
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?
What is the **purpose of the data generation or re-use** (in other words, what is the justification for the collection/processing/generation of the data, how and for which purpose will the data be collected/processed/generated/used)? How does it relate to the **objectives of TANGO** (i.e., for which WP/task/deliverable is it relevant)?
Please explain whether any **manipulations** to the data(set) will take place during the project (e.g., modifications after original creation, addition of newly collected data, anonymization, etc.).
Please explain who "owns" the dataset and who will have the rights to **control access** to the dataset.
Please indicate the partner in charge of **collecting, storing and deleting** the data.
Identify the TANGO 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.
Please identify the partner(s) **processing** the data included in this dataset and describe the activities that will be performed.
Please explain whether the dataset will be **combined** with any other datasets (from TANGO partners or third parties).
To whom may the data be useful outside of the TANGO project ("**data utility**")?
Will personal data included in your dataset be **transferred to countries outside the EU/ EEA**? If so, please specify.

2. Data storage during the project, quality and security

Please explain how and where the data will be **stored** during the project (e.g., local storage, cloud storage, storage media etc.).
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 provide an overview of the measures that you have or will put in place to ensure **data security** (including data recovery, as well as secure storage/archiving and transfer of sensitive data). Please include in your answers information about, but not necessarily limited to, physical security, network security, and security of computer systems and files.

3. Data availability and sharing between TANGO partners during the project

For datasets that are available to other TANGO 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 to personal data/confidential information).
Please provide **technical information on access to data** (e.g., tools or software needed to access and use the data), the **record-keeping** of identities, data alteration and access events, and the **protocols for access to data** (e.g., Is authentication needed? Is there a data access request procedure?)?
Please explain the mechanisms put in place to ensure that data is available when needed to authorized users for the purposes of TANGO.

4. Archiving and preservation 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. If not all data will be kept, please explain which data will be kept and how the selection of data for preservation will take place. Please also explain why some data cannot be preserved (e.g., legal or contractual restrictions, storage/budget issues, institutional policies, etc.). Please also specify in your answers for how long any back-ups will be held and how they will ultimately be destroyed.

5. FAIR data

5.1. Making data findable, including provisions for metadata

Will data be identified by a **persistent identifier**? If not, please explain why.

Will rich **metadata** will be provided to allow discovery? What metadata will be created? Will disciplinary or general **standards** be followed, and if so which? 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 optimize the possibility for discovery and potential re-use?
What **naming conventions** do you follow?
Do you provide clear version numbers?
Will metadata be offered in a way that it can be **harvested and indexed**?

5.2. Making data accessible

Data
Will all data be made **openly available**?
If certain datasets cannot be shared (or need to be shared under restrictions), please explain (i) **what restrictions** will apply (including identifying the data that will not be made available) and (ii) **why** these restrictions apply. In doing so, please clearly separate legal and contractual reasons from voluntary restrictions.
How will data be made accessible (e.g., by deposition in repository)?
Please explain **when** will data be made available to third parties (non-TANGO partners) (e.g., as soon as possible after the data collection, at the end of the project, at the time of a publication, etc.?). If an **embargo** is applied, please explain why and for how long it will apply, keeping in mind that research data should be made available as soon as possible.
What **methods or software tools** are needed to access the data?
Will the data be accessible through a free and **standardized access protocol**?
If there are restrictions on use of the data, 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)?
Metadata
Will metadata be made **openly available** and licenced under a public domain dedication CCO? If not, please explain 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 the data is no longer available? For how long will metadata remain 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)?
Repository
Where will the data and associated metadata, documentation and code be deposited? Will they be deposited in a trusted repository? If so, please indicate the repository chosen.
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?

5.3. Making data interoperable

Are the produced data interoperable, allowing data exchange and re-use between researchers, institutions, organisations, etc. (in other words, do they adhere to standards for formats, in line, as much as possible, with available (open) software applications, and do they enable re-combinations with different datasets from different origins)?
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? If so, which ones?
If it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will your provide **mappings** to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow re-using, refining or extending them?

Will your data include qualified references to other data (e.g., data from your project or datasets from previous research)?

5.4. Increasing data re-use

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 to the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use **licenses**? Please explain.
Will the data produced in the project be **useable** by third parties, in particular after the end of the project? Please provide an overview.
Will the provenance of the data be thoroughly documented using the appropriate standards?
Describe all relevant **quality assurance processes.**

6. Other research outputs

Please identify **research outputs other than data** that you will generate during the TANGO project. Such outputs may be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., samples, new materials, etc.).
Please explain in sufficient detail how these research outputs will be managed, shared and made available for reuse. In doing so, please consider the FAIR principles identified above.

7. Allocation of resources

What **costs** will you incur for making data or other research outputs FAIR (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)?
How these costs will be covered?
Who will be responsible for data management in your project?

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).

8. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

9. Other issues

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)?

Annex 2

1. Introduction

1.1. This document represents the policy of [partner's name] (hereinafter – **'we'/'us'/'Organization'**) with regards to the processing of personal data within the TANGO project.

1.2. TANGO project (Digital Technologies Acting as a Gatekeeper to information and data flows) is a research project currently run under the Horizon Europe Framework Programme under the Grant Agreement no. 101070052.

1.3. The EU-funded TANGO project through its technologies aims to design a trustworthy environment acting as a gatekeeper to information and data flows. Citizens and public/private organizations will be empowered to act and interact providing data both online and offline. TANGO will focus its activities on 3 main pillars: (i) the deployment of trustworthy, accountable and privacy-preserving data sharing technologies and platforms; (ii) the creation of data governance models and frameworks; (iii) the improvement of data availability, quality and interoperability – both in domain-specific settings and across sectors.

2. Scope of the policy

2.1. This policy only considers the processing of personal data by the Organization concerning its participation in the TANGO project. Other processing activities carried out by the Organization are outside the scope of this policy.

2.2. The participation of the Organization in the TANGO project will include [the type and description of activities, pilots where the Organization is involved, other relevant details].

2.3. The Organization's activities in TANGO will include the following processing of personal data: [types of personal data and data subjects, processing activities, and purpose(s) of processing].

2.4. Data will be processed according to "the principles relating to processing of personal data" of Article 5 GDPR, while the legal basis for processing personal data is defined by Articles 6 or 9 GDPR, depending on the personal data that will be processed and the circumstances under which the process is taking place.

3. Data protection principles

We support the principles set out by the GDPR by the following measures:

3.1. **lawfulness, fairness, and transparency of processing**: before and during the processing of personal data based on any legal basis stated in Article 6 GDPR, we provide the data subjects with information sheets describing the legal basis on which the processing is being done, the personal data that is going to be processed as relevant to the project and the processing activities. We inform them about their rights (mentioned in part 4 of this policy), providing them with all the necessary information regarding the way the subjects could proceed to any request associated with their rights and the way partners are obligated to respond.

3.2. **purpose limitation**: we only process personal data that is necessary to reach the goals of the project.

3.3. **data minimization**: we only collect and process personal data that is strictly necessary to conduct our activities in TANGO.

3.4. **accuracy of data**: we update/modify/erase the data upon request of the data subject or upon other discovery of its incorrectness.

3.5. **storage limitation**: we store the data during the term of the project and for [term] after its finishing; we irrevocably delete the data or anonymize it after the end of its processing.

3.6. **integrity and confidentiality**: we limit the scope of people having access to personal data to those who work in our organization and participate in the project; we store the personal data separately and use authentication procedures to control the access to it; [other applicable security measures to be added];

3.7. **accountability**: we use this policy to set and demonstrate compliance with the GDPR [other ways to be compliant and demonstrate it if necessary].

4. Rights of data subjects.

We respect the rights of data subjects specified in the GDPR, including:

4.1. the right to ask us what data is being collected about the data subject and how this data will be used in connection with the TANGO project ("right to access").

4.2. the right to lodge a complaint with a supervisory authority ("right to complain").

4.3. the right to request us to correct any of data subject's personal data that is inaccurate ("right to rectification").

4.4. the right to request us to erase, without undue delay, data subject's personal data ("right to erasure" also "right to be forgotten"), unless such a request would render impossible or seriously impair the achievement of the objective of that processing – including the impairment or invalidation of the research. According to Article 17(3) GDPR, such request can be denied in cases that processing is necessary: (a) for exercising the right of freedom of expression and information; (b) for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (c) for reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3); (d) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or (e) for the establishment, exercise or defence of legal claims.

4.5. the right to request us to restrict the processing of data subject's personal data ("right to restriction of processing").

4.6. the right to receive the personal data related to the data subject which he/she has provided to us and to transmit this data to another controller ("right to portability"); and

4.7. the right to object, at any time, to us regarding the processing of data subject's personal data ("right to object").

5. Other provisions

5.1. This policy is effective as of [date] till the end of the processing activities [the period after the end of the TANGO project to be defined].

5.2. This policy will be reviewed annually and updated if needed until the end of its effect.

5.3. We propose to the partners to strongly consider the appointment of a Data Protection Officer in cases they proceed to operations which require regular and systematic monitoring of data subjects on a large scale, especially if special categories of data are being processed.

5.4. All our employees having access to personal data specified herein, will be informed on this policy and other measures expected from them to be compliant with the GDPR.

5.5. The person controlling the implementation of this policy and other measures to comply with the GDPR from the side of the Organization is [name and contact details of the Organization's employee representing TANGO].

5.6. In case of questions regarding data protection rules and implementation of this policy the Organization will consult with [Name Surname (Entity Name), e-mail:]

Representative of the Organization: _____

Place / date: _____

Signature: _____