D7.2 First data management plan

Submission date: 31/03/2024

Due date: 31/03/2024

DOCUMENT SUMMARY INFORMATION

Grant Agreement No	101120763	**Acronym**	TANGO
Full Title	It takes two to tango: a synergistic approach to human-machine decision-making		
Start Date	01/10/2023	**Duration**	48 months
Project type	Research and Innovation Action (RIA)	**Funding programme**	Horizon Europe
Deliverable	D7.2: First data management plan		
Work Package	WP7 - Project management		
Type	DMP — Data Management Plan	**Dissemination Level**	PU - Public
Lead Beneficiary			
Author	Ronald Chenu Abente Acosta [UNITN]		
Reviewer	Artur Bogucki [CEPS]		

Funded by the European Union

DISCLAIMER

COPYRIGHT

DOCUMENT HISTORY

Version	Date	Changes	Contributor(s)
V0.1	26/02/2024	First draft version for internal review	Ronald Chenu Abente Acosta
V1	28/03/2024	Revised version with feedback from partners	Ronald Chenu Abente Acosta
V1.1	31/03/2024	Quality controlled version	Ronald Chenu Abente Acosta

PROJECT PARTNERS

Partner	Country	Short name
UNIVERSITA DEGLI STUDI DI TRENTO	IT	UNITN
TECHNISCHE UNIVERSITAT DARMSTADT	DE	TUDa
UNIVERSITA DI PISA	IT	UNIPI
CONSIGLIO NAZIONALE DELLE RICERCHE	IT	CNR
SCUOLA NORMALE SUPERIORE	IT	SNS
UNIVERSITE PARIS CITE	FR	UPC
FONDAZIONE BRUNO KESSLER	IT	FBK
CARR COMMUNICATIONS LIMITED	IE	CARR
ISTRAZIVACKO-RAZVOJNI INSTITUT ZA VESTACKU INTELIGENCIJU SRBIJE	RS	IVI
SURGICAL SCIENCE SWEDEN AB	SE	SuS
UNIVERSITATSKLINIKUM HEIDELBERG	DE	UKHD
CENTRE FOR EUROPEAN POLICY STUDIES	BE	CEPS
BCAM - BASQUE CENTER FOR APPLIED MATHEMATICS	ES	BCAM
U-HOPPER SRL	IT	UH
INTESA SANPAOLO SPA	IT	ISP
EIT DIGITAL	BE	EITD
SHARE FONDACIJA	RS	SHARE
A11-INITIATIVE FOR ECONOMIC AND SOCIAL RIGHTS	RS	A11
MINISTARSTVO ZA BRIGU O PORODICI I DEMOGRAFIJU	RS	MFWD
AZIENDA PROVINCIALE PER I SERVIZI SANITARI	IT	APSS
SWANSEA UNIVERSITY	UK	UoS
THE UNIVERSITY OF WARWICK	UK	UoW
MARTIN LUTHER UNIVERSITY HALLE WITTENBERG	DE	MLU

Funded by
the European Union

LIST OF ACRONYMS

Acronym	Definition
DoA	Description of Action
EC	European Commission
HaDEA	European Health and Digital Executive Agency
WP	Work Package
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
ICT	Information and Communications Technology
ODbL	Open Database License
OpenAIRE	Open Access Infrastructure for Research in Europe
POPD	Protection of Personal Data
WP	Work Package
HDSS	Hybrid Decision Support Systems

EXECUTIVE SUMMARY

This document provides the first Data Management Plan (DMP) for the TANGO project. The deliverable provides an overview of how research data that is generated or collected during and after the project will be managed, while also describing which standards will be followed and how this data will be shared.

It is important to note that, at the moment of the writing of this deliverable, the TANGO project has not yet completed its first six months. Furthermore, the tasks related to the organisation of the local pilots start only nine months into the project. This means that many of its data-management related activities have not yet started and are currently being designed.

As contemplated in the TANGO Description of Action, the DMP will be treated as a living and changing document. Three further official releases for the DMP document are contemplated in the TANGO DoA (in M18, M36 and M48), though intermediate internal releases will be considered according to the progress and needs of the project.

Please note that this document, after the introductory sections, follows the template for the Data Management Plan provided by the European Commission[1].

[1]https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Data Summary

The TANGO project objectives include the development of hybrid (AI and human) foundations, paradigms and decision support systems for real world applications. To achieve this hybrid approach the TANGO project necessarily requires the involvement of human participants and the management of their related data.

In a Work Package (WP) specific basis the following are the project's need for managing data:

- WP1: through the work in tasks T1.2 and T1.3, the TANGO work package 1 will execute experimental validation of hybrid human and ai theories. More than the standalone case studies organized by WP5, these experimental validations will take the form of small exercises and experiments with human participants. In the DoA, WP1 tasks mention that some of the datasets produced by these activities will be published in Open Data repositories so this will be further developed on this and future versions of this deliverable.
- WP5: in charge of coordinating (T5.1), organising and managing the main four TANGO case studies (T5.2, T5.3, T5.4 and T5.5). These case studies are used to evaluate and provide data for the developed TANGO AI systems. More information about these four use cases is discussed in the next sub section.
- WP2 and WP3: research-centric work packages that use original research and information coming from the case studies to propose and define AI algorithms. These include algorithms for synergistic human machine learning (WP2) and algorithms for hybrid decision-making (WP3).
- WP4: a technical and implementation focused work package. WP5 will specify and redesign (T4.1) the TANGO proposed AI systems, to then engineer and implement (T4.2) for their use in the project, which then would be able to be deployed and operated (T4.3) for their use in project demonstrators (T4.4). These activities imply the systematic management of user data and as such will be focused extensively on this and future versions of this document.
- The rest of the project: the rest of the project is organised to facilitate the organisation of the previous activities (WP7) and disseminating its key results for maximum impact (WP6). This last one point is of key importance and will be further discussed on this and future versions of this document.

1.1 Definitions on research data

The TANGO project has committed to follow an Open Research Data approach to the management of research data. The project will collect and generate data mainly related to the case studies that will be designed, deployed, and piloted within the scope of WP5. The rest of the project will use this data internally for their research-centric objectives and, finally, a carefully chosen subset/summarization of that data will be published in open repositories.

These three stages and categories for the data managed in the TANGO project have been summarised, along with their key involved elements, in the following figure.

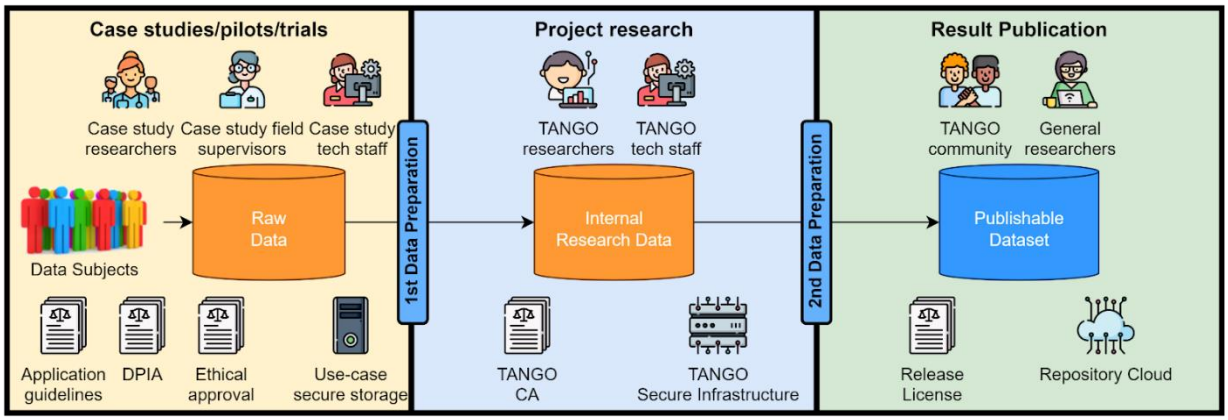

Figure 1: The TANGO project three data management categories and their involved persons, storage and regulating documentation.

1. Raw Data: data collected or generated during the TANGO project activities towards the achievements of its specified goals. This includes statistics, experiment results, measurements, and observations resulting from fieldwork, survey results, and interview recordings.
2. Internal Research Data: obtained from the previous level Raw collected/generated data by performing data preparation, anonymization and summarization operations. This category of data can be requested by TANGO partners for performing activities towards the project objectives.
3. Publishable Dataset: It is expected that the datasets produced during the TANGO project are of interest to researchers of a community wider than the TANGO project itself. Nevertheless, due to the sensitive nature of some data, only a limited subset, properly anonymised and aggregated, will be made publicly available for further research.

These categories and details for their management will be discussed throughout the rest of this document.

1.1.1. Datasets to be produced during the project

The TANGO project will need to produce and release numerous datasets of the three identified data management categories to achieve its objectives, and in this subsection the specifics related to the formats, dimensions and structure (fields and internal organisation) of this information will be specified in future versions of this document.

Currently the TANGO project is before the end of its month 6 and the tasks related to the organisation of the pilots start only in month 9, so this DMP document and D5.1 (the Trial Handbook) can only focus on general rules and practices and preliminary information at this point of time.

Generally speaking, at the current stage of the project, we can identify three main categories of data that will be generated through research activities in TANGO: personal data on the user (inc. data from personal health records in the perinatal wellbeing one), behavioural data and interactions data. To better give an idea of the context in which this will generated we present below a quick summary of the four TANGO use cases:

- Perinatal health and wellbeing: this pilot will be focused on developing a personal assistant supporting women in making decisions during pregnancy and post-partum.
    - *Participants*: MLU, TUDA, UH, UoS, UNIPI, UNITN.

- *Activities*: design, implementation, and evaluation of a digital screening and interactive agent for decision support to help early detection of diseases and inferior peripartum outcomes.
- *Location*: The study will be conducted at Depts for Halle and Tubingen in accordance with the Declaration of Helsinki, with participants recruited following defined inclusion criteria and data acquisition approved by the ethical board of Halle University.
- *Evaluation*: The study will randomly assign participants to one of two groups: an intervention group, which will engage in interaction with the TANGO agent, and a control group, which will receive usual treatment. Evaluation will include (i) usefulness, (ii) user-satisfaction, (iii) user engagement, (iv) user behaviour. The study will provide feedback for cognitive studies.

- Supporting surgical teams in making intraoperative decisions: involves complex medical scenarios and teams of expert human decision makers. The objective is developing a tool for hybrid intraoperative decision making in the surgical operating room.
  - *Participants*: UNITN, APSS, SuS, UH, UoS, UNIPI, UPC.
  - *Activities*: the trained agent will be integrated in the simulator running on the augmented operating room at DISI, and it will interactively engage with the team before and during the operation, to support them in making the most appropriate decisions.
  - *Location*: the task will be run in cooperation with the Centre for Medical Science of UNITN, APSS and the laboratory of Prof. Marco Zenati at the Harvard Medical School
  - *Evaluation*: the TANGO agent will be evaluated in simulated surgical operations where its effectiveness in contributing to the success of the endovascular surgery will be measured based on the reduction in the number of incorrect decisions made by a trainee surgical team when interacting with the agent, compared to the number of incorrect decisions made when operating without the support of the agent.
- Supporting loan officers and loan applicants in credit lending: will develop a decision support system for credit lending, helping both bank officers to make fair and safe decisions on loan applications, and applicants to receive transparent explanations and recommendations.
  - *Participants*: FBK, ISP, UNITN, SNS, UNIPI, CNR, CEPS, BCAM.
  - *Activities*: design, implementation and evaluation of an interactive TANGO agent supporting bank officers and bank customers in the credit lending decision process. The Data Science and AI unit from the partner (ISP) will provide the domain expertise and data.
  - *Location*: Italy
  - *Evaluation*: The system will be evaluated with respect to answering the following questions: does the interaction orchestrated by TANGO: 1. increase the trust of the loan officer in AI systems? 2. help the loan officer assess applicants' creditworthiness? 3. help the loan applicant understand the outcome? 4. help the applicant reach the desired outcome?
- Supporting policy makers in shaping social welfare policies: the objective is developing a tool for hybrid social policy making, aimed at ensuring fairness and transparency in government incentive allocation in the domain of family welfare and demography.
  - *Participants*: IVI, FBK, MFWD, A11, SHARE, CEPS, BCAM
  - *Activities*: design, implementation and evaluation of a TANGO Hybrid Decision Support system supporting policymakers within Serbia's ministry of family care and demography (MFWD).

Different stakeholders will interact with the TANGO framework through a web portal, connecting the AI agent, policy makers and citizens.

- *Location*: Serbia
- *Evaluation*: The system will be evaluated in terms of both policy maker and citizen satisfaction.

1.1.2. Pre existing data

As stated in the TANGO DoA, the following datasets are available at the start of the project:

- UKHD/HALLE: has 3 years of anonymised data about perinatal outcomes and patient-reported-data.
- UNITN: anonymised data consisting of around 1,000 surgical procedures performed at the Santa Chiara Hospital in Trento in the last 10 years.
- ISP: 3+ years of anonymised data about credit lending procedures and their outcomes
- MFWD: long-term data concerning compensation for vulnerable groups such as parents with children with special needs, children without parental care, financially disadvantaged families, just to name a few. MFWD also has data about the demographic picture in the context of balancing the age structure, maintaining the spatial balance of the population and the revitalization of rural areas.

These datasets will be kept in their original institution and managed under their previous data management agreements and their current licences.

2. FAIR Data

The TANGO project has committed to following FAIR data principles as defined in H2020 Programme Guidelines on FAIR Data Management[2]. This section will describe different ways that the project is preparing to follow these principles but, due to being still early in the lifetime of the project, many details related to the exact structure of data and datasets are still under consideration. This information will be refined and detailed in future versions of this document.

2.1. Making data findable

It is expected that the datasets produced during the TANGO project are of interest to researchers of a community wider than the TANGO project itself. As such, one of the main objectives of the data preparation operations for converting TANGO Internal Research Data to Publishable Data is to properly document it with machine-readable metadata to facilitate its discovery.

To facilitate finding and using of the resulting data, best practices for data stewardship (including data cleaning and curation, data preservation) will be adopted. All data will be supported by a rich set of metadata, expressing in a machine-readable format how the data was collected and processed. Finally, a rich set of keywords and a DOI will be provided for significant releases to further favour the indexing of the produced datasets.

[2] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2.2. Making data accessible

All TANGO partners are committed to making their results as open as possible, as early as possible, and as available as possible. The principles of openness and transparency will underlie all research activities, fostering sharing and collaboration as early as possible, and throughout the project´s life.

Data will be made accessible and findable also through Zenodo or similar open data repositories. The specific repository to store the TANGO project will be decided after a careful analysis. The considered factors for that analysis will encompass a number of qualitative issues that may completely override the numerical analysis (i.e. cost, capacity, expertise, investment, scalability, strategic importance).

With respect to data that is part of publications (e.g., articles, papers), they will be made available as open data. Open and transparent practices will be implemented in line with the open science policy in Horizon Europe, encouraging the use of the Open Research Europe (ORE) publishing platform, the European Open Science Cloud (EOSC) and the open repository for research objects, ensuring availability and findability.

Before data becomes of the Publishable Data category and is able to be published, and following what is stipulated in the TANGO GA and following the TANGO CA, an embargo period may be applied on the data.

2.3. Making data interoperable

The TANGO consortium in its effort to make the project's datasets as accessible and interoperable as possible will seek compliance with open standards. More specifically, data will be stored using open data formats and implementing open standards that will follow best practices and guidelines for working with open data.

Data stored will be accompanied by relevant metadata to ensure re-usability of the data by third parties. The project will adhere to W3C Data on the Web Best Practices.

2.4. Making data reusable

Datasets produced in the project will be annotated with documentation that contains, in a machine-readable way, global information about the dataset's origin (date and location of creation, details about the experiment that created it and the specific objectives for doing so) and its encoding format. Furthermore, each field/column will also be documented with information about its semantics, format, accuracy and other details. When considered necessary, a human-readable codebook will be generated for each dataset explaining details related to the execution of the activities that produced the dataset and possible issues and limitations for the dataset.

Continuing the project's commitment to Open Science, permissive and open licences will be used to maximise reusability of the produced datasets. More information will be provided in future releases of this document.

3. Other research outputs

Computer runnable code is one of the key 'other research output' identified at this stage of the project so more details are given in the subsection below. Other potential outputs (e.g., training materials, techniques, methodologies, modelling materials) along with their dissemination will be discussed in future versions of this document.

3.1. Computer Code

The TANGO project source code policy is in line with the open science philosophy. As stated in the DoA, the TANGO software library will be released as open source under an Apache v.2 licence. To maximise the impacts of the project activities, single components of the TANGO software ecosystem will be published (in the category `AI assets') in the AI-on-demand platform developed by AI4Europe.

This will ensure that the research community at large will be able to independently use the TANGO methods and algorithms, improve them and combine them to devise novel solutions to high-impact projects. At the same time, this will also allow innovators and entrepreneurs to build sustainable business propositions enabled by the TANGO results, leading to new ventures and job creation.

4. Allocation of resources

Most of the costs related to data management and GDPR/FAIR compliance have been already accounted for in the project proposal and then the TANGO DoA. The effort for these activities has already been distributed among the different data management centric work packages (mainly WP5 but also WP1, WP4, among others).

As compiled in D5.1, the trial handbook, the current responsible persons for each of the pilots is compiled in the following table.

Table 1: Trial Team roles, names and organisations.

Trial Name, Task	Trial Owner, Organization	Technical Coordinator, Organization	Practitioner coordinator, Organization	Evaluation coordinator, Organization
T5.2 (Improve/Increase) perinatal health	Prof. Dr. Markus Wallwiener (HALLE)	Tim Tobiasch (TUDa)	Prof. Dr. Stephanie Wallwiener (HALLE)	Prof. Dr. Eva Kantelhardt (HALLE)
T5.3 A surgical decision making	Andrea Passerini (UNITN)	Erich Robbi (UNITN)	Annalisa Trianni (APSS)	Stefano Bonvini (APSS)
T5.4 - Credit Lending	Bruno Lepri (FBK)	Simone Centellegher (FBK)	Daniele Regoli (ISP)	Andrea Cosentini (ISP)
T5.5 - Policy	Dubravko Culibrk (IVI)	Slobodan Ilic (IVI)	Ana Martinovic (SHARE)	Milica Marinkovic (A11)

At this stage of the project there is no clear indication of the volume or amount of data that will need to be stored and secured for achieving the project objectives. As these requirements become clear this section will be updated in future versions of this document with its details and management/security policies.

5. Data security

This section defines the general data security provisions for the two more restrictive TANGO data categories.

More concrete technical and operational details of how these requirements will be implemented have not been fixed yet, at the current state of the project when this document has been released. Nevertheless, this section will be updated with those details in future versions as the TANGO infrastructure is developed and the case studies and applications specified.

5.1 Provisions for Raw Data

The respective pilot, case study or application leaders (see Table 1) will serve as the data controller for the raw collected dated data and will be responsible to implementing the following security measures:

- Storage that is secure, encrypted and password protected by individual access credentials in the infrastructure from its institution.
- Transfers from and to this storage should be similarly encrypted and protected by individual access credentials.
- Participants will be identified by the use of pseudonymized study IDs
- The study's keys and access will be restricted only to the study/pilot staff and the developers of the applications.

5.2. Provisions for Internal Research Data

Research data is obtained from the previous level Raw collected/generated data by performing a data preparation operation. This data preparation operation will be carried out, under the responsibility of the original data controller of the raw collected data, within six months of the completion of the study that generated it and will involve:

1. The anonymization of the collected data by deleting the participant pseudonyms, other study keys and the sociodemographic/sensitive data.
2. The duly justification of any remaining personal and sensitive data to make sure that it is relevant and limited to the purposes of the project activities (in accordance with the 'data minimisation' principle).

The resulting data is called Research Data and has the following security requirements:

- Storage that is secure and password protected by individual access credentials in the infrastructure from its institution.
- Transfers from and to this storage should be similarly protected by individual access credentials.
- This information will be kept for at most 10 years after the end of the pilot / case study that generated it.

The TANGO project consortium partners are the primary target audience for Research data. To access it, the TANGO partner will send a written request to the dataset controller in which the following information will be

contained: i) name of the persons to be granted access, ii) duration of this access and iii)activities for which the data will be used. The data controller will then decide to approve the request and share access information if the request is aligned with the TANGO project objectives.

The partner receiving access to Research Data, commits to:

- Follow the rules in the TANGO CA and TANGO GA related to data management and secrecy.
- To use the data during the timeframe and for the activities specified in the written access request and to delete it after these purposes have been achieved.
- To keep data secure and comply with the same security requirements as the data controller.

6. Applicable regulations

This section will detail the legislations, regulations and guidelines that the project will consider for its data management procedures and practices.

6.1. Ethics

6.2.1. General Data protection

When personal data is involved, the TANGO project will ensure compliance with the General Data Protection Regulation which regulates the processing of personal data in the EU and its aim is to ensure individual fundamental rights. This includes the following:

- DPIA, all TANGO case studies will act according to the article 35 of the GDPR performing the data protection impact assessment when needed. This includes providing detailed information
  - On what personal data is collected, stored and processed;
  - On the recruitment process, inclusion/exclusion criteria for participation;
  - On privacy/confidentiality and the procedures that are implemented for data collection, storage, access, sharing policies, protection, retention and destruction during and after the project;
  - On how informed consent is pursued;
  - if application/is needed to be filed with a local/institutional ethics review body (if personal data is being collected) and if yes, which bodies / where / when.
- All research and development activities conducted within the scope of the TANGO project will be mindful of the “Privacy by Design” principle of article 25 GDPR, as well as article 22 GDPR assuring compliance with automated individual decision-making including profiling.
- Some of our case studies require sensitive data input (like for example pregnancy support and loan decisions) so the management of these in a responsible, ethical way is crucial for the project.

The TANGO project must process personal data under the Agreement in compliance with the applicable EU, international and national law on data protection (in particular, Regulation 2016/67916). The project must ensure that personal data is:

- processed lawfully, fairly and in a transparent manner in relation to the data subjects
- 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 they are processed
- accurate and, where necessary, kept up to date
- 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.

The TANGO project may grant its consortium members access to personal data only if it is strictly necessary for implementing, managing, and monitoring the project. The TANGO project must ensure that its members having this access are under a confidentiality obligation.

6.2.2. Raw Data management

The previously defined TANGO Raw Data category is specifically relevant as it is the one containing the most sensitive information in the project and as such should be the most protected.

The data management of this category of data will be guided and regulated by the following documents (as seen in Figure 1):

- Application Guidelines: document with guidelines for the implementation and coordination of the applications. It will provide common guidance, best practice and support for the TANGO applications and case studies. This will be released as deliverable D5.1 in M6.
- DPIA: each data processing operation where personal data is involved (like project applications or case studies) requires the creation of a Data Protection Impact Assessment (DPIA). The objective of this document is to help assess, identify, and minimise risks that may result from data processing. DPIAs are further used to interact, get feedback and ultimately the approval of privacy protecting officers and ethical committees.
- Ethics Approval: each data processing operation where personal data is involved is required to be approved by a local ethics commission. The existence of this document ensures that the project activities and their particular implementation in the target location, have been deemed ethically appropriate. The partner in charge of the application or case study must submit the request for review of their institution's ethics committee and get its approval. In the absence of the ethics committee figure, the partner must work with the coordinator to identify the alternative suitable figures to perform this task. As part of the Ethics approval, two further documents must be also created:
    - Privacy Statement: document used to inform the potential participants on the technical and organisational measures to safeguard the rights of the research participants. The main

objective of this document is to promote transparency and to give individuals more control over the way their data is collected and used.

- Informed Consent: to be processed, personal data requires free and fully informed consent from the involved persons. Obtaining in advance and clearly documenting the data subject informed consent ensures that all participation in these project activities are voluntary and the subjects are aware of their rights.

6.2.3. Proof of due diligence documentation

In order to keep an auditable document trail of the project data management activities and agreements, the following signed documents need to be accessible digitally in a project shared storage space:

- DPIAs: final version of the DPIA document, with the written approval (through email or document) by the institution data protection authority.
- Ethical Approvals: final version of the document and all associated material submitted to the Ethical Committee for their approval. The letter or email where the committee expresses their feedback and approval should also be preserved.
- Privacy Statements: the final text of the Privacy Statement given/displayed to study participants.
- Proof of informed consent: either a signed document (in the case of a paper/in person-based study) or a table summarizing the names/pseudonyms (in the case of a digital based study) of the persons that have voluntarily accepted participating in the study.

This information will be preserved for at most 10 years after the conclusion of the studies they refer to.

6.2.4. Non-personal data

In terms of the governance of non-personal data, the Non-Personal Data Regulation and Open Data Directive both will be complied with while designing and running AI applications between project partners while exchanging datasets or re-using public datasets for project purposes. Furthermore, case studies will adopt state of the art ethical standards for datasets and model reporting such as Model Cards or Datasheets frameworks. Project will also comply with national data protection legislation and relevant technology norms and standards.

6.2.5. AI, Marketplaces and Gender considerations

Beyond the Data Protection aspects and the requirements of approval by local Ethics Committees that were previously mentioned and, because of the type of objectives and subjects that the TANGO project is focused on, significant regulatory progress is expected throughout the duration of the project with regards to the topics of Digital Markets and Artificial Intelligence.

To take into the account this combination of both mature and evolving regulation and make sure that the TANGO project complies with its commitment to ethical research practices, it will follow the following four-point strategy:

1. Compliance with regulations will be achieved by considering a broad legal framework at every relevant step of the project. The machine learning development process and the case studies specifically will be continuously assessed for compliance with existing and forthcoming EU Directives and Regulations.

Such existing legal texts include in particular Directive 2000/31/EC on electronic commerce, Directive 2002/58/EC on the processing of personal data and the protection of privacy in the electronic communications sector (E-Privacy Directive), Regulation (EU) 2016/679 best known as General Data Protection Regulation (GDPR), Regulation (EU) 2018/1807 on the free flow of non-personal data, and the Directive (EU) 2019/1024 on open data and the reuse of public sector information.

2. The Ethics Guidelines for Trustworthy AI will be considered at every step of the project and provide an additional AI governance framework to ensure responsible use and creation of AI. We will ensure that the project's methodology complies with the 'do no significant harm' principle as per Article 17 of Regulation (EU) No 2020/852 on the establishment of a framework to facilitate sustainable investment, also known as the EU Taxonomy Regulation. This means that the project is designed in a way it is not significantly harming any of the six environmental objectives.
3. At the heart of the compliance assessments will be a data justice approach [Solano, et al. 2022], which centres on equity, recognition and representation of plural interests, and the creation and preservation of public goods. Beyond compliance with existing EU regulations, the TANGO coordinator will continuously monitor if a pilot or the TANGO technology requires consideration of a broader normative context. Soft law approaches like the digital rights theory, and industry standards will be consulted where law lacks sufficient effectiveness to mitigate risks related to the cutting-edge technology employed in the case studies. The project's outcomes will be future proofed through a focus on compliance with forthcoming regulations like the so-called AI Act, the AI Liability Directive, the Data Act, and the Data Governance Act.
4. Where appropriate, cross-disciplinary recommendations for future regulation will be put forward with the aim of enhancing effectiveness and efficiency of the governance of cutting-edge AI technology

Since the main legislative act tasked with fundamental rights protection in the AI domain is still a work in progress, the project will turn to soft normative frameworks for AI governance. Furthermore, the project is set to run in the interval of substantial legal volatility regarding digital economy regulation. The next years might see the normative landscape of the AI governance transformed under not only horizontal regulations such as the AI Act but also specific sectoral regulation making compliance an even more hectic endeavour than it regularly is. TANGO partners are aware of this specificity and will adjust the compliance strategy accordingly to the legal reality.

In particular, to prevent gender bias in machine learning models, the TANGO project will develop methods for mitigating unexpected, at times, catastrophic failure in learning caused by subgroup imbalance within the training data. In the context of algorithmic fairness, subgroups refer to demographic attributes including overlapping dimensions of race, sex, gender, age, and disability (intersectional fairness). Methods for auditing inappropriate bias against those protected subgroups will also be developed.

Finally, it is worth noting that the TANGO project has an ethical advisor that will closely follow and produce written assessments and feedback that will be taken into account for the evolution of the data management strategies in the project.

6.2. Legislation

All partners will adhere to relevant national and international laws, guidelines and policies, including:

1. Legislation and recommendation on human rights, dignity and integrity of the user
    a. Universal Declaration of Human Rights (United Nations)
    b. Regulation No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals
    c. Charter of Fundamental Rights of the European Union (2010/C83/02)
    d. Convention for the Protection of Human Rights and Fundamental Freedoms (Council of Europe)
2. Legislation and recommendations on personal data
    a. EU General Data Protection Regulation (GDPR) on the protection of individuals with regard to the processing of personal data and on the free movement of such data and was designed to harmonise data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organisations across the region approach data privacy. The General Data Protection Regulation No 2016/679 applies from 25 May 2018.
    b. Article 29 Working Group 05/2014 Opinion on Anonymization Techniques.
    c. Handbook on European data protection law by the European Union Agency for Fundamental Rights and the Council of Europe (2013)
    d. Serbia Data Protection Law (Official Gazette of RS 87/2018) from 9 November 2018, with its applicability starting from 21 August 2019.

7. Other Issues

7.1. Non-EU Countries

In the case personal data is transferred from the EU to a non-EU country or international organisation, it must be in accordance with Chapter V of the General Data Protection Regulation 2016/679. In case personal data is transferred from a non-EU country to the EU (or another third state), it must comply with the laws of the country in which the data was collected. Furthermore, the project provides that non-EU members will be compliant with European regulations as legally binding in the signed Grant agreement. Therefore, the enforcement on these countries will be based on the procedure and documents produced in the context of European legislation, with measures that safeguard local law.

As previously stated, the TANGO project will comply not only with the General Data Protection Regulation (GDPR) from the European Union, but also will adhere to high ethical standards for data management. Even Non-EU project members, who signed the Grant Agreement, will be legally bound to the GDPR Regulations and have committed to follow the project's ethical standards.

The challenge of transferring data from the European Union to project members in non-EU countries and vice versa will be addressed by establishing adequate technical and organisational measures protecting the personal data from our data subjects across continents and considering intercultural information ethics as the basis for data protection activities.

8. Conclusions

This document represents the first TANGO Data Management Plan (DMP) that was delivered at the end of the sixth month of the project. It is important to note that no activities related to data collection or generation have yet officially started. And as such the policies and procedures contained in this document should guide the creation of the more operative documents to be used for those activities.

Three levels of category of data within the TANGO project have been defined in this document: i) raw collected/generated data; ii) research data; and iii) publishable data. Data processing pipelines and methodologies have been established for each. It is also defined that to the extent consented by privacy directives all data that is generated will be openly available as FAIR data, using interoperable data formats and published under licences facilitating reuse. Furthermore, these datasets will be uploaded for long term preservation to the Zenodo or similar open repositories.

As the project keeps evolving and different types of activities are planned and executed, each subsequent version of the DMP will contain further development and specification of its different sections. Three further official releases for the DMP document are expected: a) at the project's month 18; b) at the project's month 36, with most the use case and applications started, this revision will focus in much more detail in the operative details related to data generation/collection and the different related data processing pipelines; c) at the project's month 48, at the project's end, the focus will be on how to licence and distribute the different results produced in the project.

ANNEXES

A1: Annex I – Partner Data Management Consultation

The information below was compiled through a consultation with all project partners and it gives a general idea of the data management roles, the category of data that they will have access to (from the three TANGO defined data categories), the general use they play to give to the data and links to additional institutional data protection references.

Table 2: TANGO partners number, short name, summary of activities and data use category.

#	Partner	Activities	Role and data use	Additional institutional data protection rules
1	**UNITN**	Project management (WP7). Leading the synergistic human-machine learning WP (WP2). Developing algorithms for shared decision making (WP3) and leading the case study on the augmented operating room (WP5).	Case study leader - surgical teams: access to internal research data and raw data only when necessary	https://www.unitn.it/en/ateneo/51105/privacy-and-data-protection
2	**TUDA**	Developing neuro-symbolic learning approaches for improved human-machine interactions via understandability and reliability (WP2).	Research partner: access to internal research data	https://www.tu-darmstadt.de/datenschutz/index.de.jsp
3	**UNIPI**	Leading the hybrid DM WP (WP3), developing algorithms for cognition-aware explainability methods for interactive ML (WP2) and contributing to their use in case studies (WP5).	Research partner: access to internal research data	https://start.unipi.it/dataprotection/dataprotection_ricerca/
3.1	**CNR**	Leading the task on cognition-aware explanations for interactive DM (WP3), contributing to the design of cognitive-aware explanations for interactive learning (WP2) and to the research for both individual decision making and Social Policy Making (WP3).	Research partner: access to internal research data	https://www.rpd.cnr.it/wp/
3.2	**SNS**	Leading the task on human-machine co-evolution (WP2), contributing to the research on cognition-aware explainability methods (WP2-3) and its use in case studies (WP5).	Research partner: access to internal research data	-
4	**UPC**	Developing a folk theory of hybrid decision making (WP1). Revising the theory based on feedback from the case studies (WP5).	Research partner: access to internal research data	The ethical comittee decides on a case by case basis

Funded by the European Union

5	**FBK**	Leading the social policy making research activities (WP3) and contributing to its corresponding case study (WP5). Leading the credit lending case study (WP5) and contributing to the exploitation of the results via its research and innovation network and joint labs with companies and PAs.	Case study leader - credit lending: access to research data and raw data only when necessary	-
6	**CARR**	Leading the impact generation activities (WP6); coordinating all dissemination and communication efforts; supporting clustering, networking and user engagement activities; ensuring clarity and consistency of the TANGO visual identity across solutions.	Impact partner: access to publishable data and internal research data if justified	-
7	**IVI**	Evaluating the potency of the TANGO framework by leading the policy-making case study (WP5), as well as devising and implementing methods for social policy making (WP3), and refining theories based on the proofs-of-concept and case studies (WP1).	Case study leader - policy making: access to research data and raw data only when necessary	-
8	**SuS**	Providing technical and software engineering skills for the surgical team decision making case study (WP5), by developing functionalities to interface its simulator for virtual reality surgical operating rooms with the TANGO API.	Research partner: access to internal research data	-
21	**MLU**	Evaluating the effectiveness of TANGO (WP5) as a supporting agent during pregnancy and post-partum. Crucially, this WP will also provide feedback for refining the social interaction theory (WP1) and the AI decision explanation theory (WP3).	Case study leader - perinatal health: access to research data and raw data only when necessary	-
10	**CEPS**	Providing an analysis of the ELSE factors that arise in theorizing, developing and applying the project to ensure ethical decision making (WP2, WP3, WP5). CEPS will also monitor compliance with existing and forthcoming EU law, and provide recommendations for policy making (WP3).	Research partner: access to internal research data	https://www.ceps.eu/about-ceps/data-privacy-policy/
11	**BCAM**	Leading the development of reliable models with fairness and bias considerations as part of WP2. Translating those reliable models into ethical and trustworthy decision making (WP3), and contributing to case studies (WP5).	Research partner: access to internal research data	https://www.bcamath.org/en/privacy-policy
12	**UH**	Leading implementation and integration activities (WP4). Design and implementation of the TANGO API. Support for development and integration of apps in WP5. Exploitation planning for the TANGO API (WP6).	Research partner: access to internal research data	-

13	**ISP**	Contributing to the design and implementation of the credit lending case study (WP5) and planning the exploitation of its outcomes.	Case study direct actor - credit lending: acces to raw data	https://www.intesasanpaolo.com/it/common/footer/privacy.html
14	**EITD**	Leading the Exploitation activities of TANGO (WP5) by coordinating the execution of the planned Pilots and supporting the identification of additional exploitation opportunities in collaboration with pilot owners.	Impact partner: access to publishable data and internal research data if justified	-
15	**SHARE**	Ensuring the involvement of citizens in all stages of design, implementation and evaluation of the HDM system in the policy-making case study (WP5).	Case study direct actor - policy making: acces to raw data	https://www.sharefoundation.info/en/privacy-policy/
16	**A11**	Involving most vulnerable citizens for the design, implementation and evaluation of the HDSS, while taking into account citizen satisfaction for the policy-making case study (WP5).	Case study direct actor - policy making: acces to raw data	-
17	**MFWD**	Supporting agent as it is involving policy makers and assisting in evaluating the system formed for the policy-making case study (WP5).	Case study direct actor - policy making: acces to raw data	https://www.paragraf.rs/propisi_download/zakon-o-finansijskoj-podrsci-porodici-sa-decom.pdf
18	**APSS**	Providing domain expertise, data and the trainee surgical team for the case study in T5.3.	Case study direct actor - surgical teams: acces to raw data	-
19	**UoS**	Developing methods to foster decision-making based on interaction, dialogue and reason, to produce model(s) for human-agent co-evolution (WP3). Ensure synergistic systems will learn through interaction design capturing user need (WP2) and drive this into deeper algorithmic advances.	Research partner: access to internal research data	https://www.swansea.ac.uk/about-us/compliance/data-protection/
20	**UoW**	Building a theoretical framework for mutual understanding using the theory of "virtual bargaining," (WP1). Revising the theory based on feedback from the case studies (WP5).	Research partner: access to internal research data	https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection/