



## DELIVERABLE REPORT D1.2

### DELIVERABLE

SUBMISSION DATE	NAME OF THE DELIVERABLE	WORK PACKAGE
NATURE	AUTHOR(S)	LEAD BENEFICIARY
28.2.2023	D1.2 Data Management Plan	WP1
Public	Massimo Magaldi	ENG

### PROJECT DETAILS

PROJECT ACRONYM	TITAN	GRANT AGREEMENT	101070658
CALL IDENTIFIER	Human-01-27	PROJECT DURATION	01.09.2022 – 30.09.2025
PROJECT OFFICER	Peter Freiss	PROJECT COORDINATOR	Engineering S.p.A.

## QUALITY CONTROL ASSESSMENT

VERSION	DATE	DESCRIPTION	NAME	ORG
v0.1	13/2/2023	Table of contents	Massimo Magaldi	ENG
V0.2	23/2/2023	First Draft version	Massimo Magaldi	ENG
V0.3	16/3/2023	Added contributions	Various contributors	ENG, DBT, 21C, VTT
V0.4	6/4/2023	Second Draft version	Massimo Magaldi	ENG
V0.5	13/04/2023	Added contribution	Marcin Woźniak	SWC
V0.6	21/04/2023	Review and Feedback	Ana Inguanzo, Trisha Meyer	VUB
V1.0	28/04/2023	Final version after revision	Massimo Magaldi	ENG

## DISCLAIMER

The opinions stated in this report reflect the opinions of the authors and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein. All intellectual property rights are owned by the TITAN consortium members and are protected by the applicable laws. Except where otherwise specified, all document contents are: “©TITAN Project - All rights reserved”. Reproduction is not authorised without prior written agreement. The commercial use of any information contained in this document may require a license from the owner of that information.

## STATEMENT OF ORIGINALITY

This Deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## ACKNOWLEDGEMENT

This document is a deliverable of the TITAN project, which has received funding from the European Union’s Horizon 2020 Programme under Grant Agreement (GA) #101070658

TABLE OF CONTENTS

**EXECUTIVE SUMMARY ..... 4**

**1 TITAN DATA MANAGEMENT GENERAL GUIDELINES ..... 5**

1.1 INTRODUCTION.....5

1.2 PERSONAL DATA HANDLING .....5

1.3 GATHERING TRAINING DATASETS .....6

1.4 FAIR DATA .....6

1.5 ETHICS.....6

    1.5.1 *Personal Data & Informed Consent* ..... 7

**2 DATA MANAGEMENT DETAILS PER PARTNER ..... 8**

2.1 21C .....8

2.2 DBT ..... 11

2.3 ENG ..... 14

2.4 NCSR-D..... 17

2.5 VTT..... 19

**3 ANNEX 1: TITAN CITIZEN CO-CREATION WORKSHOP CONSENT FORM .....25**

## EXECUTIVE SUMMARY

The Data Management Plan (this report) is a living document that presents the consortium's plan on handling data during and after the end of the project; what data will be collected, processed and/or generated, which methodology and standards will be applied, whether data will be shared or made open access and how data will be curated and preserved.

Since several TITAN partners will create/collect/handle data for different purposes, the TITAN DMP is structured in 1) General Guidelines, that should be followed by all partners unless otherwise stated in individual partner's plans, and 2) Data Management details per Partner.

**All consortium members should refer to and be compliant with this DMP.**

Data management in TITAN is respectful of and aligns with the project internal ethics protocol, according to the legal and ethical requirements identified in T2.1 and through internal self-assessment and the External Ethics Advisory Board (EEAB) review.

This document is the first early version of the TITAN Data Management Plan, and it will be constantly updated/revised during the project duration. Its first submission was planned at M6 but has been delayed at M8 to provide complete information about the data collected during the co-creation workshops.

## 1 TITAN DATA MANAGEMENT GENERAL GUIDELINES

### 1.1 Introduction

The TITAN Data Management Plan addresses data production, collection, and processing. It provides detailed information about the project data lifecycle, privacy, and the policies for data collection, storage, access, sharing, protection, retention, and destruction.

**All consortium members should refer to and be compliant with this DMP.**

TITAN will manage the research data generated in the action responsibly and in line with the FAIR principles.

This DMP is inspired by a template modelled after the guidelines specified in Science Europe's *Practical Guide to the International Alignment of Research Data Management*, aiming to ensure that the core aspects of the DMP are covered and that the data produced and used in the project are FAIR: Findable, Accessible, Interoperable and Re-Usable.

Since several TITAN partners will create/collect/handle data for different purposes, the TITAN DMP is structured in 1) General Guidelines, that should be followed by all partners unless otherwise stated in individual partner's plans, and 2) Data Management details per Partner.

### 1.2 Personal Data handling

TITAN collects personal data for two purposes:

1. to organize the co-creation & pilot activities and ensure the diversity of the recruited participants,
2. to reimburse travel expenses.

**Unless otherwise stated in individual partner's plans, Personal Data will be collected and processed based on informed consent and will be deleted (or completely anonymized for authorized sharing) upon completion of the co-creation & pilot activities.**

Indeed, participants in the TITAN project activities will only be engaged upon having given informed consent in accordance with the with GDPR (A 6.1(a)), and following the principles of informed consent, according to which consent should be freely given, specific, granular and be the result of an affirmative action on part of the data subject. This will be operationalized throughout the activities of the project using a template for obtaining consent for processing of participants' personal data. The partner responsible for any given TITAN project activity which processes personal data will provide an email address through which consent can be revoked. A consent revoke-system will be developed for the TITAN environment.

Data used to reimburse travel expenses will need to be stored for auditing purposes following Horizon Europe and national regulations.

Personal data will be particularly secured in the following ways:

- (1) only data necessary for the research will be collected, in accordance with the data minimization principle.
- (2) the data will be stored on system-encrypted servers that allows fine-grained access controls at the file-level.
- (3) If any sensitive data is stored on local devices (laptop, external HDD), these will be equipped with an up-to-date OS and antivirus.
- (4) Person-related data will be pseudonymized as there is no need for the output to be personally identifiable.
- (5) Any data security incident will be reported instantly to the partner's Data Protection Officer, the coordinator's Data Protection Officer, and any other relevant instances.

### 1.3 Gathering Training Datasets

To train the AI models supporting the TITAN intelligent coach under the hood, training datasets will be gathered during co-creation sessions (e.g. living labs) through surveys and collecting entire dialogue sessions between participants and the intelligent coach mockups.

Even no special category of personal data will be explicitly requested from living labs participants, they may express political opinions and religious or philosophical beliefs during the dialogue with the intelligent coach.

Thus, as part of the informed consent procedure and in line with TITAN's aim to create a sustainable, human centric and trustworthy AI ecosystem that is complies to European AI legal & ethical requirements, **participants will be informed about their interaction with AI technology, the abilities, limitations, risks and benefits of its use, and that entire dialogue sessions could be collected and pseudonymized.**

### 1.4 FAIR data

TITAN adopts an open science approach based on open cooperative work with citizens and systematic sharing of knowledge and tools as early and widely as possible during the project. This way, the project actively contributes to the on-going efforts towards an open, distributed, trustworthy AI environment for citizens against disinformation. TITAN's open science approach aims to accelerate the advancement of knowledge and innovation via an open, inclusive approach (by involving all relevant stakeholders including citizens and civil society) and by sharing the project results, making them more findable, accessible and reusable and improving their reproducibility.

Resulting data sets, when possible, according to partner's IP and TITAN's legal and ethical requirements, will be released under open licenses.

TITAN Parties, (i.e., legally represented partners), agree that any background, results, confidential information and/or any and all data/information that is provided, disclosed or otherwise made available between the Parties during the implementation of the Action and/or for any Exploitation activities ("Shared Information"), will not include personal data as defined by Article 2, Section (a) of the Data Protection Directive (95/46/EEC) and applicable local implementing local legislation; or, as from May, 25th 2018, Article 4 of the General Data Protection Regulation. Accordingly, each partner managing gathered data will ensure that all data and information contained in shared information is anonymised such that it is no longer personal data, prior to providing the Shared Information to such other Parties.

### 1.5 Ethics

The TITAN project will fulfil its aim of developing an ethically sound and socially responsive AI environment to fight disinformation through a human-centred co-creation approach where citizens and relevant stakeholders from several representative EU countries are engaged during workshops and other engagement activities (i.e. Living Labs).

Data management in TITAN is respectful of and aligns with the project internal ethics protocol, according to the legal and ethical requirements that are currently being set in T2.1 and taking into account the indication of the External Ethics Advisory Board (EEAB) that will guide the consortium to obtain ethics approvals for the participation of humans, to safeguard proper handling of personal data, to manage AI issues properly, and to mitigate potential risks from project's results being used by third parties to improve their misinformation methods and tools.

TITAN will assess the impact of project activities on the legal and ethical requirements identified in T2.1, through (1) iterative self-assessment, in which each partner contributing to the development of AI solutions reports on the legal and ethical risks and challenges associated with the proposed research by completing a self-assessment questionnaire, (2) an internal legal and ethical impact assessment executed by VUB and (3) external expert review conducted by the EEAB

TITAN's legal and ethical requirements will be defined in D2.1 (TITAN Socio-technical Framework and User Needs Analysis) and subsequently translated into a collaborative self-assessment questionnaire and an impact assessment template in T1.4 (Legal and Ethical Compliance).

**For more in-depth and updated details about legal and ethical issues, see the template and guidance for legal and ethical impact assessment (D1.3) and the Annual Report on legal and ethical impact assessment (D1.4).**

### *1.5.1 Personal Data & Informed Consent*

Participants in the TITAN project activities will only be engaged upon having given informed consent in accordance with the with GDPR (Art 6.1(a)), and following the principles of informed consent, according to which consent should be freely given, specific, granular and be the result of an affirmative action on part of the data subject. This will be operationalized throughout the activities of the project using specific templates for obtaining consent for processing of participants' personal data (for the Co-creation Phase A, see Annex 1: TITAN Citizen co-creation workshop consent form). As part of the informed consent procedure and in line with TITAN's aim to create a sustainable, human centric and trustworthy AI ecosystem that is complies to European AI legal & ethical requirements, participants will be informed about their interaction with an AI technology.

TITAN Parties, (i.e., legally represented partners), agree that any background, results, confidential information and/or any and all data/information that is provided, disclosed or otherwise made available between the Parties during the implementation of the Action and/or for any Exploitation activities ("Shared Information"), will not include personal data as defined by the General Data Protection Regulation. Accordingly, each Party will ensure that all data and information contained in shared information is anonymised such that it is no longer personal data, prior to providing the Shared Information to such other Parties.

## 2 DATA MANAGEMENT DETAILS PER PARTNER

The TITAN Data Management Plan addresses data production, collection, and processing. It provides detailed information about the project data lifecycle, privacy, and the policies for data collection, storage, access, sharing, protection, retention, and destruction.

This DMP is inspired by a template modelled after the guidelines specified in Science Europe’s *Practical Guide to the International Alignment of Research Data Management*, aiming to ensure that the core aspects of the DMP are covered and that the data produced and used in the project are FAIR: Findable, Accessible, Interoperable and Re-Usable.

Since several TITAN partners will create/collect/handle data for different purposes, this section is structured in Data Management details per Partner.

### 2.1 21C

1. General Information	
Partner Name	21c Consultancy
Partner Contact Details	Susie McAleer Susie@21cconsultancy.com Acorn House, 33 Churchfield Road, London W3 6AY, UK
2. Data description	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data



<p>Explain how new data will be collected or produced and/or how existing data will be re-used.</p> <p>Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).</p> <p>If you <b>reuse</b> existing data, specify the <b>source</b> of these data.</p>	<p><b>TITAN website (newsletter subscription and contact)</b>  TITAN website allows visitors to keep up to date with TITAN latest news and events. The dataset will comprise of visitors' contact details, which will be gathered via contact forms, asking users to voluntarily fill out personal details.  The dataset contains the following textual data:</p> <ul style="list-style-type: none"> <li>• Name/Surname</li> <li>• E-mail address</li> <li>• User's Occupation/Organization</li> <li>• Message by the user.</li> </ul> <p>The data are stored in database format on the hosting platform and will not be made openly available as its sole purpose is to keep track of, get in contact and work with the people registered to the different mailing lists and to the work environment of the project.</p> <p><b>TITAN Twitter and LinkedIn analytics</b>  We will collect Twitter tweets traffic (tweets, likes, retweets) and LinkedIn traffic (share link, share comment, share url, follows, connections (name, surname, company, position), invitation sent (name, date), logins (IP address), member follows, messages received (message title, sender, sender profile, date, subject, content), users (age, gender)).  The dataset is provided by Twitter and LinkedIn and collected by 21c as the TITAN Communication &amp; Dissemination lead and Social Media manager only for project internal use to gain valuable insights to improve project communications.  The dataset is accessible by 21c.</p> <p>No existing data sources can be reused for the purposes described above.</p>
<b>3. Ethical and legal issues</b>	
<p>Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.</p>	<p><input checked="" type="checkbox"/> Yes  <input type="checkbox"/> No</p> <p><b>TITAN website (newsletter subscription and contact)</b>  The only personal data that will be collected will be through the website newsletter subscription and contact. More specifically the data collected will be</p> <ul style="list-style-type: none"> <li>• Name/Surname</li> <li>• E-mail address</li> <li>• User's Occupation/Organization</li> <li>• Message by the user.</li> </ul> <p>All this information will be manually and voluntarily inserted by website visitors and will be used solely to keep track of, get in contact and work with the people registered to the different mailing lists and to the work environment of the project.</p>
<p>Are there any ethical or legal issues concerning the creation and/or use of the data?</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No</p>

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes, please comment: Twitter and LinkedIn analytics obtained through the social media platforms should not be further re-used, in line with the specific provisions of each platform's terms and conditions.
<b>4. Documentation and metadata</b>	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	N/A: the data collected in the framework of 21C's activities in the project are not to be re-used or shared with anyone.
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> N/A If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	TITAN project repository on Microsoft SharePoint Online.
How will the data be backed up?	According to the Microsoft SharePoint Online backup policy.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes Yes, the Website contacts, Twitter, LinkedIn datasets don't require a lot of storage space. <input type="checkbox"/> No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered?	These costs are covered by ENG and are part of ENG's project budget.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	The TITAN Microsoft SharePoint Online folder is accessible only by consortium members.
<b>6. Data preservation after the end of the TITAN project</b>	
Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	There's no need to retain Newsletter and website contacts, Twitter, LinkedIn datasets after the end of the project.
Where will these data be archived (= stored for the long term)?	The datasets will be deleted and not archived.
What are the expected costs for data preservation after the project? How will the costs be covered?	N/A
<b>7. Data sharing and reuse</b>	

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes If yes, please specify: Newsletter and website contacts are personal data. We gained informed consent solely for the indicated purposes and not for sharing.
Which data will be made available after the end of the project?	None
Where/how will the data be made available for reuse?	<input type="checkbox"/> In an Open Access repository <input type="checkbox"/> In a restricted access repository <input type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):
When will the data be made available?	N/A
Who will be able to access the data and under what conditions?	None
What are the expected costs for data sharing? How will these costs be covered?	N/A
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	21c <sup>1</sup>
Who will be responsible for data storage & back up during the project?	21c
Who will be responsible for ensuring data preservation and sharing?	21c
Who bears the end responsibility for updating & implementing this DMP?	21c is responsible for the data detailed in this table, and ENG.

## 2.2 DBT

<b>1. General Information</b>	
Partner Name	Danish Board of Technology (DBT)
Partner Contact Details	<b>Teknologirådet</b> <b>Arnold Nielsens Boulevard 68E, 2650 Hvidovre</b> <b><a href="mailto:Tekno@tekno.dk">Tekno@tekno.dk</a></b> <b>+45 33 32 05 03</b>
<b>2. Data description</b>	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data

<sup>1</sup> See Partner Contact Details in the General Information section.

<p>Explain how new data will be collected or produced and/or how existing data will be re-used.</p> <p>Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).</p> <p>If you <b>reuse</b> existing data, specify the <b>source</b> of these data.</p>	<p><b>Co-creation Phase A – Workshops</b></p> <p>Origin of data: Citizens signing up for the co-creation workshops via Qualtrics, EngageSuite (online platforms), email, online form, Eventbrite, Google forms.</p> <p>Data from the following countries: Belgium, Bulgaria, Denmark, Ireland, Lithuania, Slovakia, Spain, Italy</p> <p>Type of citizen data (the same for all countries): Name, email, age, occupation, education, area of living, news habits. Furthermore, for participants in the workshop: everyday life habits, news consumption, opinions on news, disinformation and AI.</p> <p>Data format: Qualtrics, EngageSuite (online platform), internal file server, MS SharePoint system, Google Sheets, Excel Sheets, Google Forms.</p> <p>Data volume – Citizens from 8 different countries: Invited directly: 910 citizens Registered for the workshops: 445 Selected for the workshops: 260 Participated in the workshops: 214</p> <p><b>Co-creation Phase B – Living Labs</b> To Be Defined</p> <p><b>Co-creation Phase C – Pilot</b> To Be Defined</p>
---	---

**3. Ethical and legal issues**

<p>Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.</p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>All participants, across all countries, have signed and handed in a consent form ensuring GDPR regulations compliance.</p> <p>Participants personal data collected: <b>Co-creation Phase A - Workshops</b> Name, email, age, occupation, education, area of living.</p> <p><b>Co-creation Phase B – Living Labs</b> To Be Defined</p> <p><b>Co-creation Phase C – Pilot</b> To Be Defined</p>
---	---

Are there any ethical or legal issues concerning the creation and/or use of the data?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No  If yes: - Reference to ethical committee approval: ...
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
<b>4. Documentation and metadata</b>	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	To be Defined
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> N/A If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	In DBT's SharePoint, with password protection.  Results data from each of the workshop managed by partners has been internally stored by each organization in the following places (in the process of planning the workshop): Excel sheets, word documents, online form, google sheets, and unless otherwise stated in individual partner's data management details, Personal Data will be deleted (or completely anonymized for authorized sharing) upon completion of the co-creation & pilot activities.
How will the data be backed up?	According to the Microsoft SharePoint Online backup policy.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered?	No costs for data storage and backup
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	DBTs strict practices of handling data guarantee GDPR compliance and data security.
<b>6. Data preservation after the end of the TITAN project</b>	

Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	All personal data and emails from citizens will be deleted after the co-creation activities.  Data required for project auditing purposes will be stored longer, specifically signed participant lists and bank details for reimbursement purposes.  All results, in anonymized format, will be retained for the period of the project.
Where will these data be archived (= stored for the long term)?	DBT SharePoint DBTs economy system (bank details).
What are the expected costs for data preservation after the project? How will the costs be covered?	No expected costs
<b>7. Data sharing and reuse</b>	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes If yes, please specify: Signed participant lists, bank details for reimbursement purposes and list of citizens wanting to get information about the development of the TITAN project will not be shared outside of DBT.
Which data will be made available after the end of the project?	Anonymized data not subjected to legal restriction from the workshops process, from all 8 countries who hosted the citizens co-creation workshops.
Where/how will the data be made available for reuse?	<input type="checkbox"/> In an Open Access repository <input type="checkbox"/> In a restricted access repository <input checked="" type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):
When will the data be made available?	Whenever needed (is available now, March 2023)
Who will be able to access the data and under what conditions?	Upon request to DBT, to TITAN partners and authorized research institutions interested in the TITAN outcomes.
What are the expected costs for data sharing? How will these costs be covered?	No expected costs
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	DBT, Project Manager Marie Hoff
Who will be responsible for data storage & back up during the project?	DBT
Who will be responsible for ensuring data preservation and sharing?	DBT TITAN Team
Who bears the end responsibility for updating & implementing this DMP?	DBT, Project Manager Marie Hoff is responsible for the data detailed in this table, and ENG

## 2.3 ENG

<b>1. General Information</b>	
Partner Name	Engineering Ingegneria Informatica

Partner Contact Details	Marco Cipolla - <a href="mailto:marco.cipolla@eng.it">marco.cipolla@eng.it</a> Massimo Magaldi – <a href="mailto:massimo.magaldi@eng.it">massimo.magaldi@eng.it</a>
<b>2. Data description</b>	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input checked="" type="checkbox"/> Reuse existing data
Explain how new data will be collected or produced and/or how existing data will be re-used.  Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).  If you <b>reuse</b> existing data, specify the <b>source</b> of these data.	<b>Training Datasets</b>  <i>Existing Training Datasets</i> Open-source datasets to be defined will be used for NLP and NLU tasks. Examples of possible public datasets for disinformation signals detections are provided in D2.1. Such datasets can be retrieved from heterogeneous sources in the web and are mainly in textual format.  <b>TITAN Platform User Account Dataset:</b> TITAN will gather user's profile data during the user sign-in phase. User's profile data will be completely defined together with TITAN User Interface specifications, and it will be in textual form.
<b>3. Ethical and legal issues</b>	
Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No  For the <b>TITAN Platform User Account Dataset:</b> Profile data for registration <ul style="list-style-type: none"> <li>• Name, Surname</li> <li>• Email</li> <li>• Nickname</li> </ul> <p>According to the informed consent that will be gained from the user, this data will be used solely for registration and login purposes.</p>
Are there any ethical or legal issues concerning the creation and/or use of the data?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
<b>4. Documentation and metadata</b>	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	Not yet defined at this stage of the project

Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> N/A (not yet defined at this stage of the project) If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	Centralized TITAN Server to be set up.
How will the data be backed up?	Relevant data will be backup with standard approaches.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes (This is a foreseen capability) <input type="checkbox"/> No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered?	These costs are covered by ENG and are part of ENG's project budget.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	Privacy by design as well as security design principles defined during the co-creation process will be applied to define and adopt proper tools and techniques to guarantee GDPR compliance and data security. Specific tools and techniques will be established once all datasets and their use will be specified.
<b>6. Data preservation after the end of the TITAN project</b>	
Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	It is foreseen to retain the datasets created for training and testing AI-based models for future research purposes.
Where will these data be archived (= stored for the long term)?	Not yet defined at this stage of the project
What are the expected costs for data preservation after the project? How will the costs be covered?	Not yet defined at this stage of the project
<b>7. Data sharing and reuse</b>	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes Ethical and privacy restrictions may apply to data that is difficult or even impossible to anonymize.
Which data will be made available after the end of the project?	To be defined. We aim to make available the datasets gathered and used for training the AI-based models. It will depend on the anonymization techniques that can be applied, once datasets are completely specified.
Where/how will the data be made available for reuse?	<input type="checkbox"/> In an Open Access repository <input checked="" type="checkbox"/> In a restricted access repository <input checked="" type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):
When will the data be made available?	Once training datasets will be completely gathered and anonymized, if possible.
Who will be able to access the data and under what conditions?	Upon request to ENG, to partners and authorized research institutions interested in the TITAN outcomes.



What are the expected costs for data sharing? How will these costs be covered?	No expected costs
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	ENG (see Partner Contact Details) for the datasets that ENG is going to share
Who will be responsible for data storage & back up during the project?	ENG is responsible for the data detailed in this table
Who will be responsible for ensuring data preservation and sharing?	ENG is responsible for the data detailed in this table
Who bears the end responsibility for updating & implementing this DMP?	ENG is responsible for the data detailed in this table

## 2.4 NCSR-D

<b>1. General Information</b>	
Partner Name	NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"
Partner Contact Details	Georgios Petasis (petasis@iit.demokritos.gr)
<b>2. Data description</b>	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input checked="" type="checkbox"/> Reuse existing data
<p>Explain how new data will be collected or produced and/or how existing data will be re-used.</p> <p>Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).</p> <p>If you <b>reuse</b> existing data, specify the <b>source</b> of these data.</p>	<p><b>Training Datasets</b></p> <p><i>Existing Training Datasets</i> Several existing datasets – detailed description in D2.1</p> <p><i>Generated Training Datasets</i> Sample dialogues for training assistant</p>
<b>3. Ethical and legal issue</b>	
Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
Are there any ethical or legal issues concerning the creation and/or use of the data?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes, please comment: All data gathered will have IP restrictions
Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
<b>4. Documentation and metadata</b>	

What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	No documentation will be provided
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No <input type="checkbox"/> N/A (not defined at this stage of the project) If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	NCSR-D data center
How will the data be backed up?	Data will be backed up with typical back up operations within NCSR-D data centre
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No  2TB allocated for the purposes of the TITAN project
What are the expected costs for data storage and backup during the project? How will these costs be covered?	No expected cost
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	Through permissions enforced by the security of the operating system (Fedora Linux)
<b>6. Data preservation after the end of the TITAN project</b>	
Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	All anonymized data will be retained for at least 3 years
Where will these data be archived (= stored for the long term)?	These data will be archived at NCSR-D data center
What are the expected costs for data preservation after the project? How will the costs be covered?	No expected costs
<b>7. Data sharing and reuse</b>	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes IP Issues.
Which data will be made available after the end of the project?	It cannot be decided at this point
Where/how will the data be made available for reuse?	<input checked="" type="checkbox"/> In an Open Access repository <input checked="" type="checkbox"/> In a restricted access repository <input checked="" type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):  It depends on the data_
When will the data be made available?	At the end of the Project
Who will be able to access the data and under what conditions?	It will be decided in the business model of TITAN

What are the expected costs for data sharing? How will these costs be covered?	It will be decided in the business model of TITAN
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	NCSR-D (see Partner Contact Details) for the datasets that NCSR-D is going to share
Who will be responsible for data storage & back up during the project?	NCSR-D for the datasets that NCSR-D is going to share
Who will be responsible for ensuring data preservation and sharing?	NCSR-D for the datasets that NCSR-D is going to share
Who bears the end responsibility for updating & implementing this DMP?	NCSR-D is responsible for the data detailed in this table, and ENG

## 2.5 VTT

<b>1. General Information</b>	
Partner Name	VTT Technical Research Centre of Finland
Partner Contact Details	Research Team Leader Ville Ollikainen <a href="mailto:ville.ollikainen@vtt.fi">ville.ollikainen@vtt.fi</a> +358400841116
<b>2. Data description</b>	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input checked="" type="checkbox"/> Reuse existing data
Explain how new data will be collected or produced and/or how existing data will be re-used.  Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).  If you <b>reuse</b> existing data, specify the <b>source</b> of these data.	<p><b>Training Datasets</b></p> <p><b>Datasets used for Propagation Impact Assessment analysis</b></p> <p><b>Existing Training Datasets</b></p> <ul style="list-style-type: none"> <li>- (2023-03-16) Reddit Full submission corpus. Accessed in March 2023 at <a href="https://www.reddit.com/r/datasets/wiki/testing/#wiki_full_submission_corpus">https://www.reddit.com/r/datasets/wiki/testing/#wiki_full_submission_corpus</a></li> <li>- (2023-03-16) Reddit All comments. Accessed in March 2023 at <a href="https://www.reddit.com/r/datasets/wiki/testing/#wiki_all_comments">https://www.reddit.com/r/datasets/wiki/testing/#wiki_all_comments</a></li> <li>- Usage terms for both of Reddit datasets: must reference publication <a href="https://doi.org/10.48550/arXiv.2001.08435">https://doi.org/10.48550/arXiv.2001.08435</a> and credit pushshift.io</li> </ul> <p><b>Generated Training Datasets</b></p> <p>(2023-03-16) Planned: Developing a runtime data collection component for TITAN platform, storing data that relates to the elements of Existing Training Datasets, for their applicability validation.</p>
<b>3. Ethical and legal issue</b>	

Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No <i>According to the knowledge on 2023-03-16, subject to change if needed.</i>
Are there any ethical or legal issues concerning the creation and/or use of the data?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:
Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No Usage terms of Existing datasets, however, are potentially subject to change.
<b>4. Documentation and metadata</b>	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	Data model and format documentation
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> N/A (not defined at this stage of the project) If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	During project: on researchers' laptops with full disk encryption and security practices according to VTT policies. When published: To be defined
How will the data be backed up?	Local backup on Veracrypt encrypted USB drive.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes Local laptop and USB drive storages <input type="checkbox"/> No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered?	None
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	VTT security policies; external USB drives use Veracrypt.
<b>6. Data preservation after the end of the TITAN project</b>	
Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	Published open data set(s), to be defined later.
Where will these data be archived (= stored for the long term)?	To be defined.

What are the expected costs for data preservation after the project? How will the costs be covered?	To be defined.
<b>7. Data sharing and reuse</b>	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes <ul style="list-style-type: none"> <li>Data (to be defined) may contain information, with sharing prohibited.</li> <li>Usage terms of Existing datasets are subject to change</li> </ul>
Which data will be made available after the end of the project?	<ul style="list-style-type: none"> <li>References to Existing datasets</li> <li>Generated dataset used for Existing datasets applicability validation, details to be defined.</li> </ul>
Where/how will the data be made available for reuse?	<input checked="" type="checkbox"/> In an Open Access repository <input type="checkbox"/> In a restricted access repository <input type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):
When will the data be made available?	By the end of the project.
Who will be able to access the data and under what conditions?	Open datasets, if applicable.
What are the expected costs for data sharing? How will these costs be covered?	To be defined; expect negligible costs.
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	Ville Ollikainen, project manager at VTT.
Who will be responsible for data storage & back up during the project?	Ville Ollikainen, project manager at VTT.
Who will be responsible for ensuring data preservation and sharing?	Ville Ollikainen, project manager at VTT.
Who bears the end responsibility for updating & implementing this DMP?	Ville Ollikainen, project manager at VTT, and ENG

## 2.6 SWC

<b>1. General Information</b>	
Partner Name	Swarmcheck sp. z o. o.
Partner Contact Details	Research Team Leader Marcin Woźniak m.wozniak@swarmcheck.ai +48509985068  Chief Technical Officer Łukasz Wilisowski l.wilisowski@swarmcheck.ai +48500701828
<b>2. Data description</b>	

<p>Will you generate/collect new data and/or make use of existing data?</p>	<input checked="" type="checkbox"/> Generate new data <input checked="" type="checkbox"/> Reuse existing data
<p>Explain how new data will be collected or produced and/or how existing data will be re-used.</p> <p>Describe the origin, type, format of the data (per dataset). Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).</p> <p>If you <b>reuse</b> existing data, specify the <b>source</b> of these data.</p>	<p>Anonymous arguments (claims and their relationships) in natural language shared with TITAN platform with API with Swarmcheck database stored in Mongo DB. Ever node can be re-used many times. Some of the &gt; 30, 000 claims and arguments stored in this database created before the project can be reused in TITAN platform. That would be only case of reuse of existing data.</p> <p>Data structure:</p> <p>Nodes:</p> <p>Id – unique id of the thesis</p> <p>Name – content of the thesis</p> <p>Types: thesis node/source node</p> <p>When created - date</p> <p>When updated - date</p> <p>Language – language of the thesis</p> <p>Status – for eg. “approved”</p> <p>Weight – based on relationship with other nodes</p> <p>Organization Id – designated storage space for TITAN project</p> <p>Edges:</p> <p>Id: - unique id of the edge</p> <p>Source – node</p> <p>Target – node</p> <p>Type: agree, disagree, steelman, source</p> <p>When created - date</p> <p>When updated - date</p> <p>designated storage space for TITAN project</p>
<p>3. Ethical and legal issue</p>	
<p>Will you process personal data? If so, shortly describe the kind of personal data you will use and how compliance with GDPR on personal will be ensured.</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>Are there any ethical or legal issues concerning the creation and/or use of the data?</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:

Do existing 3 <sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<b>4. Documentation and metadata</b>	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	Data model and format documentation
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input type="checkbox"/> Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> N/A (not defined at this stage of the project) If yes, please specify:
<b>5. Data storage &amp; backup during the TITAN project</b>	
Where will the data be stored?	Cloud storage MongoDB
How will the data be backed up?	Daily automatic backup
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	<input checked="" type="checkbox"/> Yes cloud service provides enough storage & backup capacity <input type="checkbox"/> No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered?	To be defined.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	SWC security policies, controlled organizational access only to designated developers managed 2 step verification with Microsoft Authenticator
<b>6. Data preservation after the end of the TITAN project</b>	
Which data will be retained after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	Claims and relationships generated by collective argumentation of users for further use of the TITAN system app
Where will these data be archived (= stored for the long term)?	MongoDB cloud storage - To be defined by the end of the project.
What are the expected costs for data preservation after the project? How will the costs be covered?	To be defined.
<b>7. Data sharing and reuse</b>	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 <sup>rd</sup> party, legal restrictions)?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
Which data will be made available after the end of the project?	All argumentation generated during the project.
Where/how will the data be made available for reuse?	<input checked="" type="checkbox"/> In a restricted access repository <input type="checkbox"/> In an Open Access repository <input type="checkbox"/> Upon request by mail <input type="checkbox"/> Other (specify):

When will the data be made available?	By the end of the project.
Who will be able to access the data and under what conditions?	TITAN app users through API
What are the expected costs for data sharing? How will these costs be covered?	To be defined; expect negligible costs.
<b>8. Responsibilities</b>	
Who will be responsible for the data documentation & metadata?	Łukasz Wilisowski, CTO at Swarmcheck.
Who will be responsible for data storage & back up during the project?	Łukasz Wilisowski, CTO at Swarmcheck.
Who will be responsible for ensuring data preservation and sharing?	Łukasz Wilisowski, CTO at Swarmcheck.
Who bears the end responsibility for updating & implementing this DMP?	Łukasz Wilisowski, CTO at Swarmcheck.



### 3 ANNEX 1: TITAN CITIZEN CO-CREATION WORKSHOP CONSENT FORM

## Consent form TITAN Citizen Co-creation workshop

---

Dear participant,

Thank you for taking part in this workshop.

In order to organize the event, we will collect information about your:

- Name
- Age
- Gender
- Area of living
- Education
- Email
- [payment information and address, for transport reimbursement]

We will collect this data to ensure diversity in the people participating in the workshop and to be able to contact you with practical information.

Your personal data will only be used for this purpose and will not be shared with any third parties.

Your contributions during the event will be anonymous and will after the event be used for analysis in a public report.

Information with personal data used to organize the meetings will be kept until all administration has been finalized, except for the sign-up sheet and payment information, which we are legally obliged to store for financial reporting, for 5 years after the project is completed in 2026. Then it will be deleted.

The workshop is organized by [please fill in your own name and organisation], on behalf of The Danish Board of Technology and TITAN. The data protection team for the event is Francisca Nordfalk and Nicklas Bang Bådum of The Danish Board of Technology (email: [gdpr@tekno.dk](mailto:gdpr@tekno.dk)).

You can withdraw your consent to participate at any time and request that your data be deleted by contacting [please fill in your responsible project manager, organisation and her/his email]. However, you cannot withdraw contributions when they are anonymized, or data necessary for financial reporting.

I have understood the terms and would like to participate.

Date: \_\_\_\_\_

Signature: \_\_\_\_\_