

KT4D

Knowledge Technologies
for Democracy

Project Number: 101094302

Start Date of Project: 01/02/2023

Duration: 36 months

Deliverable D1.1

Data Ethics Review and DMP

Dissemination Level	PU
Due Date of Deliverable	31/07/2023, Month 6
Actual Submission Date	31/07/2023
Work Package	WP 1 Project Management and Coordination
Task	T1.4 Data Management Plan and Ethics Oversight
Type	Report
Version	V.1
Number of Pages	p.1 – p. 56

Deliverable Abstract

This Data Management Plan (DMP) takes into account best practices around the management of research data as will be found in the KT4D project, including consideration of FAIR and open research data, as well as, specific privacy, data protection and security standards with respect to the data processing activities of the consortium. The aim of this DMP is to ensure maximally beneficial, but also lawful, secure, and ethically sound data processing, sharing, and reuse in line with both current and proposed EU legislation in the context of the project's activities.



KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

The information in this document reflects only the author’s views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided “as is” without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



DELIVERY SLIP

	Name	Partner/Activity	Date
Moderated by:	Kate Francis	ICTLC	28/07/2023
	Wendy Kuyoh	ICTLC	28/07/2023
	Eva Power	TCD	28/07/2023
Approved by:	Jennifer Edmond	TCD	31/07/2023

DOCUMENT LOG

Issue	Date	Comment	Author
V.0.1	12/04/2023	Initial framework draft	Jennifer Edmond (TCD)
V.0.2	31/05/2023	Integration of new sections on data protection principles in the framework draft	Kate Franics & Wendy Kuyoh (ICTLC)
V.0.3	19/06/2023	Revision of framework draft	Jennifer Edmond (TCD)
V.0.4	10/07/2023	Integration of data protection aspects of the project, data mapping tables, and overview of the draft AI Act	Kate Franics & Wendy Kuyoh (ICTLC)
V.0.5	15/07/2023	Formatting	Eva Power (TCD)
V.0.6	18/07/2023	Peer review	Keith Hyams (UW)
V.0.7	19/07/2023	Peer review	Joanne Ahearn (TRUST IT)
V.0.8	26/07/2023	Overall review of draft document	Jennifer Edmond (TCD)
V.0.9	28/07/2023	Final formatting	Eva Power (TCD)

TERMINOLOGY

Terminology/Acronym	Definition
AI	Artificial Intelligence
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
EDPS	European Data Protection Supervisor
EEA	European Economic Area
EU	European Union
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

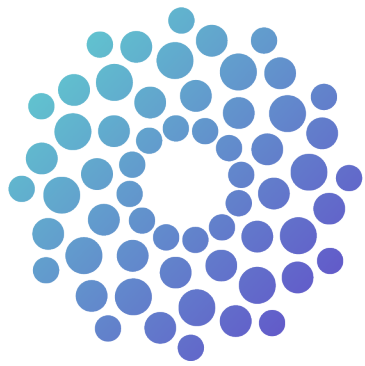
AI Act	Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence) and Amending Certain Union Legislative Acts
EC	European Commission
e-Privacy Directive	ePrivacy directive (2002/58/EFC, as revised by 2009/136/EC)
MFA	Multi-Factor Authentication
ML	Machine learning

1 Introduction	2
1.1 Purpose, Context and Scope of the Deliverable.....	2
Table 1 – Overview of WPS	3
1.2 Project summary	3
2 Detailed overview of the data types within the project (by project partner)	4
3 Legal framework	16
3.1 Processing personal data for research purposes	16
3.2 Pseudonymisation vs. anonymisation	17
3.3 EU General Data Protection Regulation (GDPR).....	18
3.3.1 Processing of special categories for research purposes under the GDPR	19
3.3.2 Lawfulness, fairness and transparency.....	22
3.3.3 Information/transparency	22
3.3.4 Fairness	24
3.3.5 Relevance and data minimisation.....	24
3.3.6 Data protection by design and by default	24
3.3.7 Rights of data subjects.....	25
3.3.8 Data Transfers.....	25
3.3.9 Profiling.....	26
3.3.10 Purpose limitation	26
3.3.11 Accuracy.....	26
3.3.12 Security and confidentiality	27
3.3.13 Storage limitation	27
3.3.14 Transmitting personal data to third parties.....	27
3.3.15 Anonymisation and pseudonymisation	28
3.4 Draft Artificial Intelligence Act	28
3.4.1 General classifications of AI systems	29
3.4.2 Obligations on the providers of high risk systems	29
3.4.3 Requirements of training machine learning models under the Act.....	29
3.4.4 Prohibited AI practices.....	29
3.4.5 Specific requirements for AI systems	30
3.4.5.1 Establishment of a risk management system.....	30
3.4.5.2 Data and data governance	30

3.4.5.3 A record of technical documentation.....	30
3.4.5.4 Record keeping.....	31
3.4.5.5 Transparency and provision of information to users	31
3.4.5.6 Human oversight	31
3.4.5.7 Accuracy, robustness and cybersecurity	32
3.4.5.8 Quality management system	32
3.5 ePrivacy Directive.....	32
4 FAIR DATA	33
4.1 Findable.....	33
4.1.1 Data sources the project builds on	33
4.1.2 Data repositories, hosting services	33
4.1.3 File naming conventions	34
4.2 Accessible	34
4.2.1 Open Access, open data policy	34
4.2.2 Documentation	34
4.3 Interoperable	35
4.3.1 Technical and social interoperability	35
4.3.2 Legal interoperability	35
4.4 Reusable	35
4.4.1 Upstream reuse - obtaining reuse rights	36
4.4.2 Downstream reuse.....	36
4.4.3 Clear licensing policy.....	36
4.4.4 Sustainability plan for the project outputs	36
5 Allocation of resources.....	36
5.1 DMP development timeline and responsibilities to update it.....	36
5.2 Resources allocated in project budget	37
6 Data security.....	37
7 Ethical aspects	37
8 Conclusion and future work	38
9 References	39
Appendix I. KT4D Record of Processing Activities v1.0.....	40
Appendix II. Personal Data Management Guidelines for KT4D Partners	53

List of Tables

Table 1 – Overview of WPs	3
Table 2 – User data	5
Table 3 – Research data	10
Table 4 – Project administration data	12
Table 5 – Dissemination and communication data	14
Table 6 – Software data	15
Table 7 – Partner personal data mapping instructions	38
Table 8 – KT4D partner personal data form	39
Table 9 – KT4D partner research data form	39
Table 10 – Beyond the Horizon (BtH) data processing activities	40
Table 11 – Fundacion Cibervoluntarios data processing activities	41
Table 12 – Demos Research Institute data processing activities	42
Table 13 – Demos Research Institute research data	42
Table 14 – Democratic Society data processing activities	43
Table 15 – Hybrid Core data processing activities	44
Table 16 – Institute of Urban and Regional Development data processing activities	45
Table 17 – Institute of Urban and Regional Development research data	46
Table 18 – Strane Innovation (STRANE) data processing activities	47
Table 19 – Strane Innovation (STRANE) research data	47
Table 20 – Trinity College Dublin data processing activities	48
Table 21 – Trinity College Dublin research data	48
Table 22 – Trust-IT data processing activities	49
Table 23 – University of Warwick research data	50
Table 24 – ICT Legal Consulting data processing activities	51



KT4D

Knowledge Technologies
for Democracy

Executive Summary

This deliverable is the first version of the Knowledge Technologies for Democracy (KT4D) Data Management Plan (DMP). With the aim of facilitating good data management practices, this DMP details the lifecycle of data collection and processing in the context of the KT4D project. In order to do so, it provides an outline of the internal processes that will be adopted in terms of data protection and security and describes the types of data that the KT4D project partners envision processing. This DMP also describes the processing, storage, and possible re-use of the data sets within the context of the project, to ensure that the research underway in KT4D will be accessible and understood by the parties involved. In this way, it contributes to ensuring the research follows the FAIR data principles.

This first version of the DMP represents a view of the consortium at this point in time (M6 of the project) and reflects the data concerned by the project at its initial stage, which is still limited. For this reason, and in order to ensure lawfulness of processing and compliance with best data protection and security practices, the DMP will be consistently enhanced and updated according to the progress of KT4D and its various work packages and tasks with further updates due at M18 and M36. The DMP will furthermore be utilised so as to achieve the project objectives in a way that is compliant with applicable EU data protection and AI-related laws.

The first part will outline the purpose, context, and scope of the deliverable. The second part will provide an overview of the types of data that will be processed by the partners throughout the project. The third part of the DMP includes a highly relevant legal framework, including the draft AI Act, data protection and data security by design, and ethical aspects related to the project. It also positions the deliverable in terms of the KT4D project's objectives. The fourth section describes the provisions made or envisioned to accommodate the FAIR principles and thereby future usage of KT4D data resources. The fifth part will consider the allocation of resources and more specifically, how the costs of making data compliant with the FAIR principles will be covered and the resources for long term preservation of data. Lastly, the DMP will set out the measures in place to guarantee data security including data recovery in the sixth part of this DMP.



Funded by the
European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

1 Introduction

1.1 Purpose, Context and Scope of the Deliverable

The purpose of this document is to ensure maximally beneficial, but also lawful, secure, and ethically sound data processing, sharing, and reuse in line with both current and proposed EU legislation in the context of the project’s activities. This is primarily guided by the guidance set out by the European Commission to manage research data according to the FAIR principles, with the aim of improving Europe’s overall provision of open data according to the principle of being ‘as open as possible and as closed as necessary.’¹ It also responds to the requirement to act within legal frameworks regarding the management of the personal data of human subjects as per the General Data Protection Regulation,² Artificial Intelligence Act.³

This DMP is a living document, which will be regularly reviewed and updated throughout the lifecycle of the project to ensure constant oversight of data processing activities. Additionally, central to the DMP is the objective of ensuring that all tools and environments the project uses to process, present, and analyse data, are sources of transformations that may have both legal implications and implications for the FAIR-ness (that ism that research data be managed in a fashion so as to make it durably Findable, Accessible, Interoperable and Reusable) of research data.

In order to achieve its goal of guiding the management of data and data systems in the KT4D project, this Deliverable contains an overview of the types of data collected within the project; a set of commitments at the project level to how FAIRness of the research data will be achieved and maintained (including data security), and a description of the full legal and ethical framework being taken into account in the development of these principles will be referenced.

Although the concept of a Data Management Plan comes with a primary focus on data and metadata by definition, the KT4D team is well aware that it cannot be comprehensive without a proper documentation and management plan of the software environment in which data is shaped, run and contextualised. Therefore, to maximise the potential of the present document we will also include software and tools developed within the project as they come online.

The different kinds of research data within the project can be categorised as follows:

OVERVIEW OF WPS:

WPs coming with admin/support data, including personal data	WP1, WP2
---	----------

¹ European Commission Directorate-General for Research & Innovation, ‘H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020’ Version 3.0 of 26 July 2016 <https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf> accessed on 10 July 2023.

² Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1.

³ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence) and Amending Certain Union Legislative Acts COM [2021] 206 final.

collected for the purposes of project dissemination	
Research data heavy WPs, collecting data according to social science methods (interview, survey, experimentation)	WP3, WP 5
Work Packages with only minimal (meta)data collection proposed	WP4, WP5, WP6
Work Packages developing relevant research software applications	WP7

Table 1 Overview of WPS

In addition to the general characterisation of the work packages given above, all work packages also contribute to the development of the KT4D Use Cases, which will collect input from user communities to shape and valourise the KT4D research outputs.

1.2 Project summary

AI and big data are fundamentally interwoven into our societies, culture and indeed into our expectations and conceptions of democratic governance and exchange. They can also, however, contribute to an environment for citizens that is distinctly anti-democratic. KT4D will harness the benefits of an understanding of these as knowledge technologies to foster more inclusive civic participation in democracy. To achieve this, we will develop and validate tools, guidelines and a Digital Democracy Lab demonstrators platform. These results will be validated across three user needs scenarios:

- 1) building capacity for citizens and citizen-facing Civil Society Organisations (CSOs);
- 2) creating regulatory tools and services for Policy and CSOs; and
- 3) improving awareness of how to design ethically and mindfully for democracy principles in academic and industrial software development.

The project’s work is underpinned by the understanding that to fully address the social and fundamental rights costs of AI and big data, we need more than just technological fixes, we need more than just technological fixes, we need to address the underlying cultural influences and barriers. Most importantly, we understand the threats to democracy of AI and big data not only through the nature of what they do, but via the cultural disruptions they create with power dynamics they shift, their tendency toward opacity, and the speed at which they change. KT4D’s ambitious and disruptive results will drive transformation in how democracy and civic participation are facilitated in the face of rapidly changing knowledge technologies, enabling actors across society to capitalise on the many benefits these technologies can bring in terms of community empowerment, social integration, individual agency, and trust in both institutions and technological instruments, while confidently mitigating potential ethical, legal and cultural risks.

2 Detailed overview of the data types within the project (by project partner)

This section provides an overview of the types of data processed within the project. A detailed overview of the data processed in the project is necessary for the purposes of ensuring that adequate safeguards are implemented by the project partners and to allow for the FAIR data principles to be followed. This section comprises five tables which provide information on User data (Table 1), Research data (Table 2), Project administration data (Table 3), Dissemination and communication data (Table 4), and Software data (Table 5). Both personal and non-personal data are included in the tables below. Further details on the personal data processed within the project are available in Annex 1, the KT4D Record of Processing Activities, Version 1 in which the envisioned processing activities of the project are detailed in accordance with Article 30 GDPR.

Data controller (Work package leader) / KT4D Partner	Data type (i.e., personal or non-personal data)	Purpose of Processing	WP/Task	Size, volume, format	Storage (location) and retention periods	Legal basis	Security measures applied and access controls
Hybrid Core (HyB)	The user-provided personal data (e.g., names, email addresses, occupations) will be processed. HyB will use the collected and tagged data in order to train the Machine Learning algorithm for	Building and deploying technical components to support the Digital Democracy Lab Demonstrator system.	WP6 (Task 6.2), WP7 (Task 7.2 and Task 7.3)	.pdf, .txt, .mpg	Data will be stored solely for the purposes of the KT4D project activities under WPs 6 and 7 and will be purged no later than three months after an event, with the exception of anonymised data required for reporting or project delivery. Data will be stored on a shared drive by the DemSoc team in a	Consent	The access is restricted to those who strictly require access to the data and will be stored in a GDPR compliant cloud servers for ML training purposes only.

	Democracy Lab.				GDPR compliant and secure location based on DemSoc’s institutional data storage practices.		
--	----------------	--	--	--	--	--	--

Table 2 – User data

Data controller / KT4D Partner	Data type (i.e., personal or non-personal data)	Purpose of Processing	WP/Task	Size, volume, format	Storage (location) and retention periods	Legal basis	Security measures applied and access controls
Fundacion Cibervoluntarios (CIB)	Name and surname, age, gender, and opinions towards political participation.	Co-creation workshops, group interviews, tool testing and validation.	Tasks 6.1 and 6.3	Stored on paper-based copies and scanned digital format, audio recordings.	Data will be stored for the shortest time possible after the end of the project. Data will be deleted after a maximum of 5 years after the end of the project. Paper based copies will be stored in CIB offices and digital copies will be stored on CIB servers.	Consent	Data will be stored in a safe server and folder within the organisation’s offices.
Demos Research Institute (DEMOS)	Name, surname, organisation, professional title and email	Use case workshops & Delphi study	Use case 3 / Tasks 4.3 and T5.1	Digital .doc.	On a shared drive by the Demos team in GDPR compliant and	Consent	Password protection.

					secure location based on the Demos data storage practices. The data will be deleted as soon as possible after the project's ending date.		
Demos Research Institute (DEMOS)	(No personal data – only publicly available author names) openly available documents such as EC communications, legislations, statements, meeting minutes and research reports	Literature review and document analysis in each task. A corpus will be developed on ethics and AI from the perspectives of e.g. global and regional governance, EU digital policy, self-assessment guidelines.	Tasks 4.1, T4.3, T5.1	.pdf, .txt. Documents will be organised by file name and made searchable for the KT4D consortium members.	Zotero will be used for storage of the organised sources lists and possibly made public at the end of the project. Data will be stored for up to five years after the project finishes.	N/A	Password protection
University of Warwick (UW)	(No personal data – only publicly available author names) Published articles and books.	For carrying out research activities and literature reviews.	Tasks 4.1, 4.2, 4.3	.pdf and .docx formats (corpora)	Zotero library set up for the project. Data will be stored for up to five years after the project finishes.	N/A	Password protection

<p>Trinity College Dublin (TCD)</p>	<p>(No personal data – only publicly available information) Research data on history of knowledge technologies, human-centred and participatory design, critical digital literacy, Participatory Algorithmic Accountability from scholarly journals and publications, conference proceedings, name, surname, websites, online resources.</p>	<p>For research purposes.</p>	<p>Tasks 3.1, 5.2, 6.1, 7.1</p>	<p>.pdf and .docx formats (corpora), and in .xlsx</p>	<p>Zotero library set up for the project.</p>	<p>N/A</p>	<p>Password protection</p>
-------------------------------------	--	-------------------------------	---------------------------------	---	---	------------	----------------------------

<p>Institute of Urban and Regional Development (IRMiR)</p>	<p>Name, gender, age and other relevant background including ideas, beliefs and preferences, email address.</p>	<p>Research (Qualitative research methods including focus groups and games). More specifically, this includes interaction design for Participatory Design Session, Use Cases 2 and 3. Focus group research, social games</p>	<p>WP3</p>	<p>.pdf, .doc, .txt. Documents will be organised by file name,</p>	<p>Data will be stored on a password protected hard drive and on the institutional cloud drive. Data will be stored during the project lifecycle and five more years upon project conclusion, as requested from the European Commission.</p>	<p>Consent</p>	<p>Password protection</p>
<p>Beyond the Horizon (BtH)</p>	<p>Name, surname and email address, age range, job affiliation or job role</p>	<p>Research and statistical reasons, specifically for events management, including Use Case consultation and co-creation meetings.</p>	<p>WP2, T2.3</p>	<p>To be determined by the partner.</p>	<p>Institutional cloud storage provided by AWS. Data will be stored during the project lifecycle and five more years upon project conclusion, as requested from the European Commission.</p>	<p>Consent</p>	<p>Access Control Policies, Data Privacy Policies, Technical Security Measures, encryption of personal data is encrypted both in transit and at rest, network Security, Data Backup and Disaster Recovery, Logging and Monitoring suspicious, Physical Security measures.</p>

STRANE (STRANE)	Innovation	Prolific IDs (or other similar research platforms), age category, level of education, socio-professional category, political orientation, behavioural data (e.g., response time, accuracy rates, etc.) and self-report data (questions about participants' attitudes, beliefs, preferences, etc.).	Research purposes, online experimental studies including the investigation of the links between prior beliefs and attitudes toward AI and big data.	T3.2 & T3.3	Qualtrics and Google Form and stored using .csv and .txt formats. Data of a maximum of 5,000 participants will be collected.	The anonymised data sets will be stored and be made open access on the Open Science Framework platform (https://osf.io/) The dataset will be stored for up to five years.	Consent	Anonymisation
Democratic (DemSoc)	Society	Information (also invented) provided by research participants in the context of Digital Democracy Lab workshops, group	Digital Democracy Lab workshops, group interviews and ethnographic documentation and observation.	WP7 (specifically task 7.2 and 7.3)	.pdf, .txt.	Data will be deleted after a maximum of 3 months after an event has taken place – with DemSoc retaining only anonymized data that is needed for	Consent	The access is restricted to those who strictly require access to the data. The data will be password protected.

	<p>interviews and ethnographic documentation and observation.</p>				<p>reporting or project delivery.</p> <p>On a shared drive by the DemSoc team in GDPR compliant and secure location based on DemSoc's data storage practices.</p> <p>The consent forms will either be digital or printed e.g., at the event itself, the paper forms will be stored on only the Demsoc's Dropbox account, the original version shall be destroyed once the scanned version is made.</p>		
--	---	--	--	--	--	--	--

Table 3 – Research data

Data controller / KT4D Partner	Data type (i.e., personal or non-personal data)	Purpose of Processing	WP/Task	Size, volume, format	Storage (location) and retention periods	Legal basis	Security measures applied and access controls
ICT Legal Consulting (ICTLC)	Name, surname, e-mail address, job title of project partners	To ensure lawful and ethical data processing activities in the context of the data processing activities	WP1, WP2, WP4	.pdf, .doc, e-mail	Data will be stored on Microsoft365 and Tresorit.	Execution of the consortium agreement	Email accounts are protected by password and MFA as is Tresorit.
Trinity College Dublin (TCD)	Name, surname, email and organisation for each member of the consortium partners who are involved with or working on the KT4D project	Project meetings and documentation: setting up mailing lists and scheduling virtual meetings (e.g., Zoom)	T1.1	.pdf and .docx formats (corpora), and .xlsx for the list of sources (bibliographies)	Data will be stored on Google Drive. The contact details will be kept for one year after the duration of the project. If a member of a Consortium partners leaves their organisation, their email will be removed from the applicable mailing lists and their access to KT4D's Google Drive will be revoked. The project also makes use of Zoom.	Execution of the consortium agreement for the performance of the project and each consortium member	Private passwords are used by each member to gain access to their Google Drive accounts. Access controls are implemented.
Beyond the Horizon (BtH)	Participant name, surname and email	Events management, including Use Case consultation	WP2	.xls, .csv	Data will be stored in a cloud platform, provided by AWS, located in Europe. Data will be stored during the project lifecycle and	Consent	Access Control Policies, Data Privacy Policies, Technical Security Measures,

		and co-creation meetings			five more years upon project conclusion, as requested from the European Commission.		encryption of personal data is encrypted both in transit and at rest, network Security, Data Backup and Disaster Recovery, Logging and Monitoring suspicious, Physical Security measures.
--	--	--------------------------	--	--	---	--	---

Table 4 – Project administration data

Data controller / KT4D Partner	Data type (i.e., personal or non-personal data)	Purpose of Processing	WP/ Task	Size, volume, format	Storage (location) and retention periods	Legal basis	Security measures applied and access controls
TRUST IT	Cookie ID, IP address	Website management	WP2	Rich Text Format (.rtf) plain text, ASCII (.txt) Hypertext Markup Language (.html)	All data retrieved from kt4democracy.eu is stored in secure data centres located in the EU. Data will be stored during the project lifecycle and for five upon project conclusion, as requested from the EC. After the conclusion of the project data will be anonymised. Retention period: from 6 months to 2 years	Consent and Performance of the contract	Access to the platform is all via https which creates an encrypted connection between the user and the website.
TRUST IT	Name, surname, email	Newsletter	WP2	To be determined by the partner.	All data retrieved from kt4democracy.eu is stored in secure data centres located in the	Consent	Organisational Security Measures, Access Control Policies, Strict access

	address, country, age, photos, job affiliation, job role, educational background, field of expertise				EU. Data will be stored during the project lifecycle and upon project conclusion, as requested from the EC. After the conclusion of the project data will be anonymised.		control policies will be enforced to ensure that only authorised personnel can access the personal data located in our Cloud provider, Data Privacy Policies, Technical Security Measures, Encryption: Personal data is encrypted both in transit and at rest, Network Security, Data Backup and Disaster Recovery, Logging and Monitoring, Physical Security measures, including access controls.
TRUST IT	Name, surname, email address, country, age, photos, job affiliation, job role, educational background, field of expertise	Webinars and Digital events registrations	WP2	To be determined by the partner.	Data will be stored in a cloud platform, provided by AWS, located in Europe. Data will be stored during the project lifecycle and for five upon project conclusion, as requested from the EC. After the conclusion of the project data will be anonymised.	Consent	Organisational Security Measures, Access Control Policies, Strict access control policies will be enforced to ensure that only authorised personnel can access the personal data located in our Cloud provider, Data Privacy Policies, Technical Security Measures, Encryption: Personal data is encrypted both in transit and at rest, Network Security, Data Backup and Disaster Recovery, Logging and

							Monitoring, Security, including access controls.	Physical measures, including access controls.
--	--	--	--	--	--	--	--	---

Table 5 – Dissemination and communication data

Data controller / KT4D Partner	Data type (i.e., personal or non-personal data)	Purpose of Processing	WP/Task	Size, volume, format	Storage (location) and retention periods	Legal basis	Security measures applied and access controls
Hybrid Core	Use of data sets licensed under CC-0, CC-by or other (still to be determined)	To build the Digital Democracy Lab Demonstrator system which will support ethical user profiling. By creating a data cooperative, the platform can ensure that users have control over their data and that it is used in a transparent and accountable way, users to contribute to the development of machine learning	WP6 (Task 6.2), WP7 (Tasks 7.2 and 7.3)	To be determined by the partner.	On a shared drive by the DemSoc team in GDPR compliant and secure location based on DemSoc's data storage Practices, Data will be stored solely for the purposes of the KT4D project activities and will be purged no later than three months after an event, with the exception of anonymized data required for	Performance of the contract	The access will be restricted to those who strictly require access to the data and will be stored in GDPR compliant cloud servers for ML training purposes only.

		<p>models and improve their accuracy and fairness, promoting ethical user profiling, by providing a sandbox environment, users can explore different profiling techniques and understand how they work without risking any negative consequence, creating user profiles while respecting privacy and confidentiality, the use of big data techniques such as usage mining and clustering can help to improve the accuracy and relevance of these profiles.</p>		<p>reporting or project delivery.</p>		
--	--	--	--	---------------------------------------	--	--

Table 6 – Software data

3 Legal framework

This section provides an overview of the applicable legal framework which is relevant for the KT4D project. Firstly, it presents brief considerations on the processing of personal data⁴ for research purposes. Secondly, it deals with the legal data protection principles outlined in the GDPR which should be complied with throughout the project by the partners. Thirdly, it deals with the draft AI Act and presents considerations which are relevant for the project. Although the following sections provide detail about the basis for decision-making about personal data in the project, we have also developed a set of project level guidelines for handling personal data which are provided in Appendix 2 below.

3.1 Processing personal data for research purposes

The objectives of EU data protection legislation such as the GDPR, dealt with in the following subsection, are to allow for the free flow of personal data within the EU, to put individuals in control of their information, and to ensure that fundamental rights and freedoms are not negatively impacted by data processing activities. The GDPR provides what the European Data Protection Supervisor (EDPS) has called a ‘special regime’ for research activities.⁵ This ‘special regime’ consists of specific derogations that organisations may rely on and of the provisions of Article 89 GDPR which require appropriate safeguards to be implemented by researchers.⁶ The special regime established entails that data protection principles (dealt with later in this section) must be complied with and that risks of research are managed in a responsible manner in compliance with the principle of accountability⁷ which requires that organisations are able to demonstrate their compliance.⁸

Specific safeguards which must be implemented in the context of research activities include appropriate technical and organisational measures, compliance with the principle of data minimisation (dealt with

⁴ Regulation (EU) 2016/679 defines personal data in Art. 4(1) GDPR as ‘any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person’.

⁵ European Data Protection Supervisor, ‘A Preliminary Opinion on data protection and scientific research’, (6 January 2020) p. 5

⁶ Regulation (EU) 2016/679, Art. 89 on ‘Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes’ reads as follows:

1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.
2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
3. Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
4. Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs.

⁷ Regulation (EU) 2016/679, Art. 5(2).

⁸ European Data Protection Supervisor, ‘A Preliminary Opinion on data protection and scientific research’, (6 January 2020) p. 2.

below), and the use of pseudonymisation when possible.⁹ Furthermore, where possible, when the purposes of the research ‘can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects’, relevant techniques should be applied to ensure that data is processed in that way which entails fewer risks for individuals,¹⁰ suggesting that anonymisation is a good technique to be adopted.

It is relevant to note that the GDPR also applies to personal data which is publicly available – i.e., tweets, Reddit posts, etc.¹¹ For this reason, due consideration must be made when making use of publicly available datasets to ensure compliance with the applicable rules. The EDPS has noted that ‘Any limitations to fundamental rights in law are to be interpreted restrictively and cannot be abused’, suggesting that the retention of personal data for indefinite periods of time by researchers or failing to adequately inform data subjects about the processing of their personal data may violate such rights.¹²

3.2 Pseudonymisation vs. anonymisation

Pseudonymisation is a technique that can often be useful to reduce risks related to data processing activities and can help organisations acting as data controllers¹³ and processors¹⁴ to fulfil their legal obligations when processing personal data.¹⁵ For the purposes of the KT4D project, it is relevant to note that the GDPR applies to pseudonymised data as it is still possible to identify data subjects when combined with other information. Pseudonymisation is defined in Article 4(5) GDPR as:

‘The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person’.

Instead, anonymised data can no longer be associated with specific people, meaning that individuals are not identifiable and the GDPR would no longer apply as the data would no longer be of the personal type.¹⁶ Recital 26 GDPR defines anonymous information as information ‘which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable’. Anonymisation, however, is particularly challenging as it is difficult to ascertain the absolute effectiveness of various techniques¹⁷ to ensure that the individuals to whom the data

⁹ Regulation (EU) 2016/679, Art. 89(1).

¹⁰ Ibid, Art. 89.

¹¹ European Data Protection Supervisor, ‘A Preliminary Opinion on data protection and scientific research’, (6 January 2020) p. 18.

¹² Ibid.

¹³ Regulation (EU) 2016/679 defines ‘controller’ in Art. 4(7) as ‘the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law’. See section 3.3 below.

¹⁴ Regulation (EU) 2016/679 defines ‘processor’ in Art. 4(8) as ‘a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller’.

¹⁵ Regulation (EU) 2016/679, Recital 28.

¹⁶ European Data Protection Supervisor, *AEPD-EDPS joint paper on 10 misunderstandings related to anonymisation*, 27 April 2021, p. 3

¹⁷ Ibid.

belongs are adequately protected; technological developments and the proliferation of data may also reduce the effectiveness of anonymisation techniques in the future.¹⁸ It may also not always be possible to reduce re-identification risks while at the same time retaining the value of the dataset for a specific objective.¹⁹

The Article 29 Working Party (the pre-GDPR group of EU data protection authorities, now known as the European Data Protection Board) has suggested, the most appropriate anonymisation solution to be used should be determined on a case-by-case basis.²⁰ It is necessary that the KT4D project partners take due care to ensure that effective pseudonymisation and/or anonymisation techniques are used when processing the personal data of research participants so as to avoid negatively impacting their rights and freedoms as a result of data processing activities in the KT4D project. Further details on the specific methods used by the project partners to pseudonymise and anonymise data will be explored in the next version of the DMP.

3.3 EU General Data Protection Regulation (GDPR)

The GDPR is considered to be the world's most comprehensive data protection law. It applies in situations where data processing activities take place using the information which relates to an identified or identifiable individual (data subject)²¹ located within the European Union (EU). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier, such as name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.²²

The GDPR deals with two types of personal data, what is referred to as **'common' personal data** and what is referred to as **'special categories' of personal data**.²³ Special category data is more sensitive and requires a higher level of protection than common personal data. Common personal data includes name, surname, nickname, address, email address, ID card number, location data, advertising identifiers, information about life events, financial information including account information, information about personal ownership or possessions, transaction information (i.e., credit, purchase, and spending habits, etc.), online identifiers for profiling such as IP address, cookie identifiers, preferences such as interests, likes, dislikes, opinions, etc. Special category data includes genetic data, health data, biometric data, information about race or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership and information on sex life or sexual orientation. In the context of the KT4D project, it may be the case that political opinions and philosophical beliefs of research participants are processed by project partners, which would thus entail the processing of special category data.

¹⁸ Ibid, p. 4

¹⁹ Ibid.

²⁰ Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques', Adopted on 10 April 2014, 0829/14/EN WP216 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed on 10 July 2023, p. 3

²¹ Regulation (EU) 2016/679, Art. 4(1).

²² Ibid.

²³ Regulation (EU) 2016/679, Art. 9.

The GDPR provides for a general prohibition of the processing of special category data. Article 9(2) GDPR,²⁴ however, provides for exemptions to such prohibition which are dealt with in the following subsection.

3.3.1 Processing of special categories for research purposes under the GDPR

While the GDPR generally prohibits the processing of special category data, as defined above, due to their level of sensitivity and potential to present certain risks to individuals, the law also provides for exemptions. In order to rely on an exemption, the organisation in charge of the data processing activity is required to apply specific measures to safeguard the data subjects whose special category data is being processed.²⁵ One exception to the prohibition to process special category data is where data processing is necessary for statistical purposes and for scientific research purposes.²⁶

Other requirements under the GDPR must, however, be taken into consideration. For example, Article 9(2)(j) GDPR stresses the necessity to safeguard the rights of data subjects. Additionally, the purpose of the data processing must be proportionate to the aim pursued in the processing.²⁷ The KT4D project will both generate

²⁴ Regulation (EU) 2016/679, Art. 9 on processing of special categories of personal data reads as follows:

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.
2. Paragraph 1 shall not apply if one of the following applies:
 - a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;
 - b) processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorised by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject;
 - c) processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent;
 - d) processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or to former members of the body or to persons who have regular contact with it in connection with its purposes and that the personal data are not disclosed outside that body without the consent of the data subjects;
 - e) processing relates to personal data which are manifestly made public by the data subject;
 - f) processing is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity;
 - g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;
 - h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;
 - i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy;
 - j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

²⁵ Regulation (EU) 2016/679, Art. 89(1).

²⁶ Ibid, Art. 9(1).

²⁷ Ibid.

data through interviews and workshops with data subjects and also potentially further process personal data to train its software. The nature of this data must be considered in order to definitively ascertain if special categories of data will be processed in order to ensure that the processing is carried out lawfully.

When processing special category personal data in the context of scientific research, in addition to the scientific research exemption, it is also possible for data controllers to rely on the explicit consent of the data subject under Article 9(2)(a) GDPR.²⁸ Indeed, the EDPS has affirmed that special category data cannot be processed in the absence of explicit consent.²⁹ Additionally, special category data may be processed when they have been manifestly made public by the data subject.³⁰

However, in relation to the last point on data made public by the subject, European data protection authorities have suggested that such provision be ‘interpreted to imply that the data subject was aware that the respective data will be publicly available which means to everyone’ and furthermore that ‘In case of doubt, a narrow interpretation should be applied, as the assumption is that the data subject has voluntarily given up the special protection for sensitive data by making them available to the public including authorities’.³¹ According to this logic, a careful case-by-case analysis must be carried out as, for example, it may be debatable if an individual would reasonably expect their social media post to be used by researchers.

³²

Article 4(11) in the GDPR defines consent as ‘any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her’. It is relevant to note that such consent under the GDPR may differ from the definition of consent in the context of, e.g., clinical trials.³³ In order to be considered as freely given under the GDPR, the data subject must be in complete control and there should not be a power imbalance between the data subject and the data controller.³⁴ Interestingly, with respect to clinical trials, the European Data Protection Board – the group of EU data protection authorities – has ‘stated that the validity of consent as a legal basis could be in doubt where a participant is in a poor condition of health or belongs to a socioeconomically disadvantaged group’.³⁵

Consent must also be specific, informed, and unambiguous. This means that the specific purposes of the data processing activity should be made clear to the research participants. Each purpose should be made known to the data subject and they should be able to consent to each purpose. However, it is often the case in research that so-called ‘broad consent’ is relied upon in research.³⁶ Recital 33 GDPR notes that:

‘It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their

²⁸ Regulation (EU) 2016/679, Art. 9(2)(a).

²⁹ European Data Protection Supervisor, *A Preliminary Opinion on data protection and scientific research*, 6 January 2020. p 19.

³⁰ Ibid.

³¹ Ibid.

³² Ibid.

³³ Ibid. p. 18.

³⁴ Ibid.

³⁵ Ibid.

³⁶ Ibid.

consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.’

The KT4D partners, when relying on consent, will comply with the provisions of Articles 4(11), 6(1)(a), 7 and 9(2)(a) GDPR and, note that as the EDPS has pointed out, ‘Specific consent normally required under the GDPR may become less appropriate in the case of collected and inferred data and especially in the case of special categories of data on which much scientific research relies’.³⁷ Furthermore, the KT4D partners must ‘carefully evaluate the rights of the data subject, the sensitivity of the data, the nature and purpose of the research and the relevant ethical standards’³⁸ and take all necessary actions to make sure that compliance with the principle of transparency (dealt with below) is achieved. When consent is relied upon, it is fundamental that such consent can be revoked at any time as easily as it was provided.³⁹

In the context of the KT4D project, personal data processed,⁴⁰ as is seen in Section 2, the main personal data processed for project management will consist of the names and email addresses of the project partners as well as data collected from research participants. Particular attention will need to be paid to ensuring that the personal data of research participants – which may include physical characteristics, pseudonyms, political opinions, and beliefs – are processed with due care and in compliance with the applicable legal framework so as to avoid negatively impacting their rights and freedoms.

The GDPR applies to data controllers⁴¹ – entities that determine the ‘why and the how’ of data processing activities, and data processors⁴² – entities that process data on behalf of the data controller according to their detailed instructions. In the context of the project, the different project partners often act as independent data controllers meaning that they will determine the purpose and the means of data processing activities necessary to reach the objectives of the tasks they are charged with carrying out in the Grant Agreement. In other cases, they may also act as joint controllers – if two or more controllers ‘jointly determine the purposes and means of processing’,⁴³ in which case the parties must determine their responsibilities for compliance with the GDPR including, e.g., with respect to transparency, and that importantly, a joint controllership agreement must be put in place between the parties. Legal partner ICTLC is currently in the process of determining how such processing relationships and roles should best be regulated between the partners. In some cases, such as the case of partner ICT Legal Consulting, however, the partners act as data processors

³⁷ Ibid. p. 19.

³⁸ Ibid.

³⁹ Ibid.

⁴⁰ Regulation (EU) 2016/679 defines ‘Processing’ in Article 4(2) GDPR as ‘any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction’.

⁴¹ Regulation (EU) 2016/679 defines ‘controller’ in Art. 4(7) as ‘the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law’.

⁴² Regulation (EU) 2016/679 defines ‘processor’ in Art. 4(8) as ‘a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller’.

⁴³ Regulation (EU) 2016/679, Art. 26.

merely processing data on behalf of, e.g., TCD, for the purpose of complying with legal data-related obligations.

3.3.2 Lawfulness, fairness, and transparency

The processing of personal data for various processing operations must include a legal basis pursuant to Article 6 of the GDPR. At least one valid lawful basis must exist for the lawful processing of personal data which must be determined prior to the processing of personal data. Legal bases under Article 6 GDPR include consent (Article 6(1)(a) GDPR), contractual necessity, meaning that the ‘processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract’ (Article 6(1)(b) GDPR), for compliance with a legal obligation (Article 6(1)(c) GDPR), to protect the vital interests of a data subject or other person (Article 6(1)(d) GDPR), for the performance of a task in the public interests (Article 6(1)(e) GDPR), and for the purpose of legitimate interest (Article 6(1)(f) GDPR).

Transparency in processing involves being clear, open, and honest with data subjects about who you are, what you intend to do with the personal data and why you need to process the specific personal data prior to the processing operations. Concretely, this is accomplished through information which is provided to data subjects when data is directly collected from them (Article 13 GDPR), when it is not collected directly from them (Article 14 GDPR), and in other cases such as in the context of a data breach which may adversely affect their rights (Article 34 GDPR).

Fairness, in contrast to transparency and lawfulness, implies that the processing of personal data shall be done in a way in which the data subjects may reasonably expect their data to be processed.⁴⁴ Furthermore, complying with the principle of fairness established under Article 5(1)(a) GDPR means avoiding any processing activities that could adversely affect the rights and freedoms of the data subject.

3.3.3 Information/transparency

Article 12 of the GDPR requires data controllers to provide information on data processing activities to the concerned data subjects via Article 13 and 14 GDPR information notices. The same article also requires controllers to inform data subjects of their rights and to do so ‘in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child’.⁴⁵ Under Article 12 of the GDPR, such information must be provided in writing or other means where appropriate such as electronic or oral means with certain caveats. In order to truly be transparent, it is therefore necessary that information provided to research participants is drafted in a clear way that is appropriate to their age, education level, level of vulnerability, etc. and that it is ascertained that they are fully aware not only of their rights but also of any potential risks that the data processing operation carried out under the KT4D project may bring about.

In terms of the information notice to be provided, on the one hand, Article 13 of the GDPR requires that where personal data is collected from the data subject, the data controller shall provide the data subjects at

⁴⁴ Information Commissioner’s Office, ‘Principle (a): Lawfulness, fairness and transparency’, < <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/the-principles/lawfulness-fairness-and-transparency/#:~:text=At%20a%20glance&text=You%20must%20use%20personal%20data,will%20use%20their%20personal%20data> > accessed 10 July 2023.

⁴⁵ Regulation (EU) 2016/679, Art. 12(1).

the time of collection, with information on the identity and contact details of the data controller and of the data protection officer, information on the purposes of the processing and the legal basis relied upon, the recipients of the personal data, if any, the purposes of processing, any potential transfers of data, the storage period of the data to be collected, information on the rights which data subjects are entitled to avail themselves of, and in cases where automated decision-making⁴⁶ may be utilised, meaningful information about the logic involved is provided, etc.⁴⁷ Within the KT4D project, it will be necessary for data subjects involved in the research to be adequately informed of their rights via a detailed and fully comprehensible information notice. The KT4D project has already drafted an information notice to be provided to research participants which can be used as a template by the various partners which will directly interact with data subjects to carry out the research activities necessary to achieve the purposes of the project.

Article 14 of the GDPR requires the data controller to provide the data subject with information where the personal data has not been obtained from the data subject, but from another source.⁴⁸ For example, in the case of research in contexts such as the KT4D project, data may not be collected directly from the data subject, but from another source. Indeed, data may be processed for a purpose which is not the original purpose for which they were collected. This is called a ‘secondary purpose’ and is compatible with the applicable legal framework so long as such secondary purpose is compatible with the original purpose.⁴⁹ Such compatibility must be carefully assessed. Concretely, in the KT4D project, it may be the case that the project partners make use of data collected from other sources which will be used to train its tools. As soon as a decision is taken on the data sets to be used, relevant data protection obligations will be identified by partner ICTLC together with the relevant data controller and this DMP will be updated accordingly.

⁴⁶ Regulation (EU) 2016/679, Art. 22, dealing with Automated individual decision-making, including profiling, which reads as follows:

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision:
 - a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
 - b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
 - c) is based on the data subject's explicit consent.
3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

⁴⁷ Regulation (EU) 2016/679, Art. 13.

⁴⁸ Regulation (EU) 2016/679, Art. 14.

⁴⁹ Regulation (EU) 2016/679, Art. 6(4) states that ‘Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia:

- (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;
- (b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller;
- (c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10;
- (d) the possible consequences of the intended further processing for data subjects;
- (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.’

The Article 29 Working Party Guidelines on transparency under Regulation 2016/679 explain that data controllers must also comply with the accountability principle under the GDPR. It follows from Article 5(2) of the GDPR that the data controller must be able to demonstrate that personal data is processed in a transparent manner in relation to the data subject.⁵⁰ Essentially, the accountability of the data controller does not only apply to the time of collection but also throughout the life cycle of data processing. This will also apply, for example, in the event when the contents of the already existing privacy notices are amended. The relevant data controllers involved in the KT4D project ought to apply the same principles when informing data subjects of the initial privacy notice and any subsequent changes to it.

3.3.4 Fairness

The fairness principle requires that personal data is processed in a manner that the data subject would reasonably expect. Fairness also entails that data is not used in a manner that would have unjustified effects on data subjects, essentially that such processing is ‘fair’. For example, the processing of personal data using AI in a manner that leads to (or, more insidiously, risks leading to) unjust discrimination is a violation of the fairness principle. The members of the Consortium aim to guarantee the individual’s rights and freedoms with regards to both their information rights and the right to not be discriminated against as a result of the data processing activities inherent to the project.

3.3.5 Relevance and data minimisation

The principle of data minimisation established under Article 5(1)(c) of the GDPR requires the data controller to limit personal data to what is directly relevant and necessary in order to achieve a specific purpose for which the data are processed. This principle will be implemented in the KT4D project through the collection and subsequent processing of personal data that is strictly necessary throughout the lifecycle of the project. Concretely, this means that only data which must be collected will be collected by the partners and any data which is inadvertently collected that is not necessary is deleted.

3.3.6 Data protection by design and by default

The principles of data protection by design and default established under Article 25 of the GDPR require the data controller, at the time of determination of the means for processing and during the processing itself, to implement appropriate technical and organisational measures which are designed to implement data protection principles in an effective manner in order to meet the requirements of the GDPR.⁵¹ Article 25 further requires the controller to implement appropriate technical and organisational measures to ensure that, by default, only personal data which is necessary for a specific purpose is processed.⁵² The principle has been implemented so far and will be implemented throughout the life cycle of the project, for instance, through the implementation of access control policies and will be given due consideration in the development of KT4D tools. The strict access control policies adopted by the Consortium partners are enforced to ensure that only authorised personnel can access the personal data located in the partner’s cloud provider. This includes the use of strong authentication mechanisms, such as multi-factor authentication (MFA), and role-based access control to limit privileges.

⁵⁰ Article 29 Data Protection Working Party, ‘Guidelines on transparency under Regulation 2016/679’ 22 August 2018, wp260rev.01 <<https://ec.europa.eu/newsroom/article29/items/622227>> accessed on 10 July 2023, p. 5, para 2.

⁵¹ Regulation (EU) 2016/679, Art. 25(1).

⁵² Ibid, 25(2).

3.3.7 Rights of data subjects

The GDPR recognises the rights of data subjects to include the right to be informed under Articles 13 and 14 of the GDPR as discussed above, the right of access to their personal data (Article 15 GDPR), the right to rectification of inaccurate information about them (Article 16 GDPR), the right to erasure or the right to be forgotten (Article 17 GDPR), the right to restriction of processing (Article 18 GDPR), the right to data portability (Article 20 GDPR), the right to object to the processing of their personal data (Article 21 GDPR), and the right of data subjects to not be subjected to decisions based solely on automated processing including profiling which produces legal or similar effects (Article 22 GDPR). The members of the consortium will fulfil the rights of data subjects upon their requests where the legal basis of processing allows for the fulfilment of the specific right.

3.3.8 Data Transfers

The ICO defines a data transfer as the intentional sending of personal data to another party or making the data accessible by it, where neither sender nor recipient is a data subject.⁵³ The GDPR has established a restriction to transfer personal data outside the European Union (EEA). More specifically, Article 44 of the GDPR establishes a general condition to comply with the provisions of Chapter V of the GDPR where the transfer of personal data is made to a third country or to an international organisation.⁵⁴ Recital 101 of the GDPR recognizes the necessity of the flow of personal data to and from countries outside the Union and international organisations as being necessary for the expansion of international trade and cooperation.⁵⁵ It further states that such transfers can take place only if, subject to conditions of transfer as laid down in the GDPR.

The GDPR does not contain a criteria for what amounts to a cross-border transmission, however, the threshold of a cross-border transfer was formulated by the EDPB in its Guidelines on the interplay between Article 3 and Chapter V of the GDPR provides a common understanding of what constitutes an international data transfer to include the following:⁵⁶

- A controller or processor (data exporter) is subject to the GDPR for the given processing. This qualification does not only apply to organisations established in the EEA but also includes organisations which, by virtue of the activities they undertake or the individuals they target, fall within the scope of the GDPR.
- The exporter discloses by transmission or otherwise makes personal data, subject to this processing, available to another controller, joint controller or processor (data importer). The EDPB provided further clarification on this qualification and noted that, it does not apply where individuals (in the course of their personal activities), on their own initiative, disclose their data directly to an organisation outside the EEA and the parties between which the disclosure takes place are not separate.
- The importer is in a third country, irrespective of whether or not this importer is subject to the GDPR for the given processing in accordance with Article 3, or is an international organisation. This

⁵³ Information Commissioner's Office (ICO) Guidelines on international data transfers, 13 July 2023.

⁵⁴ Regulation (EU) 2016/679, Art. 44.

⁵⁵ Ibid, Recital 101.

⁵⁶ European Data Protection Board (EDPB) Guidelines 05/2021 on the interplay between the application of Article 3 and the provisions of international transfers as per Chapter V of the GDPR.

means that the importer must be geographically located in a country outside the EEA, regardless of whether or not it may fall under the territorial scope of the GDPR.⁵⁷

If all the criteria established by the EDPB has been met, then the transfer constitutes a transfer to a third country or an international organisation. As a consequence, the data exporter needs to comply with the provisions of Chapter V of the GDPR and must be accompanied with one of the instruments of protecting personal data when a transfer is made to a third country or international organisation.⁵⁸ The instruments of transfer include the transfer on the basis of an adequacy decision by the European Commission recognizing the existence of an adequate level of protection pursuant to Article 45 of the GDPR.⁵⁹ Some of the other transfer instruments listed under Article 46 include: Standard Contractual Clauses (SCCs), Codes of Conduct⁶⁰ and Binding Corporate Rules (BCRs).⁶¹

As data controllers, the members of the consortium are responsible for ensuring that they comply with relevant provisions of the GDPR in the execution of their work. To this end, a checklist has been devised which will be included in the Data Management Plan.

3.3.9 Profiling

Article 4(4) GDPR defines profiling as any form of automated processing of personal data which consists of the use of personal data to evaluate certain personal aspects relating to a natural person, and in particular, to analyse or predict aspects which concern their performance at work, economic situation, health, interests, personal preferences, reliability, behaviour, location or movements.⁶² Profiling is a common data processing activity which may adversely affect the rights and freedoms of data subjects and is something which the project itself aims to explore in its research activities, i.e., in WP4 which will specifically examine how profiling may negatively impact freedom of speech.

3.3.10 Purpose limitation

The purpose limitation principle under Article 5(1)(b) of the GDPR requires personal data to be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with the original purposes.⁶³ The members of the Consortium will apply this principle by only further processing data for purposes which are compatible with the original purposes.

3.3.11 Accuracy

The principle of accuracy under the GDPR requires that personal data is accurate and kept up to date. This places an obligation to rectify or erase inaccurate data without delay.⁶⁴ Article 18(1) of the GDPR gives a data subject the right to obtain from the data controller restriction of processing where the accuracy of the personal data is contested by the data subject, for a period which enables the data controller to verify the accuracy of the personal data.⁶⁵ The members of the Consortium will proceed to promptly update any

⁵⁷ Ibid, pg 7.

⁵⁸ Regulation (EU) 2016/679, Chapter V.

⁵⁹ Ibid, Art. 45.

⁶⁰ See the EDPB Guidelines 04/2021 on Codes of Conduct as tools for transfers.

⁶¹ Regulation (EU) 2016/679, Art. 46.

⁶² Ibid, 4(4).

⁶³ Ibid, 5(1)(b).

⁶⁴ Ibid, 5(1)(d).

⁶⁵ Ibid, 18(1)(a).

incorrect information and place a restriction on the processing of personal data where information about a data subject is inaccurate when such rights are exercised by data subjects. For example, where the members of the Consortium change their email addresses, the other partners will update the mailing list to adjust to the corresponding changes.

3.3.12 Security and confidentiality

Article 32 GDPR necessitates that appropriate technical and organisational measures are implemented to ensure a level of security which is appropriate to the risk which arises as a result of the data processing activity.⁶⁶ Some of the security measures to be adopted by members of the Consortium include the pseudonymisation and encryption of personal data, especially for what concerns data that will be collected in the context of participatory research. The documentation of the project, meeting documentation to include the agendas and minutes are shared to the members of the Consortium using a Google Drive folder which has reduced the email sharing of documents. Access controls have been implemented and the drive is only accessible by the Consortium team members. The drive is also password protected. Each KT4D partner processing personal data in the context of their research activities will implement security measures which include multi-factor authentication, passwords, access controls, and where physical documents are processed, locks and keys will be used to ensure that the personal data processed are secure.

3.3.13 Storage limitation

The storage limitation principle established under Article 5(1)(e) GDPR requires that personal data is only kept in a form that allows for the identification of data subjects for as long as is necessary for the purposes of processing of personal data. This requires that time limits are established by the controller for erasure and periodic review.⁶⁷ The Consortium Partners have established a retention period which is consistent with the GDPR practices. More specifically, data will be stored for as long as is necessary depending on the circumstances as better specified in Section 2.

3.3.14 Transmitting personal data to third parties

The GDPR defines a third party as ‘a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data’. The GDPR also places a requirement for data processors to act only on documented instructions from the data controller.⁶⁸ This imposes the necessity to enter into a contract or other legal requirements pursuant to Article 28 of the GDPR with the data processor which will regulate the personal data processing relations between the data controller and third parties. The members of the Consortium will document their relations with third parties, e.g., platforms where data collected in the context of the project are stored by partners, and define their scope of obligations using a documented data processing agreement as per the requirements of the GDPR. When regulating relationships with third parties, it is necessary to make sure that eventual transfers of personal data outside the EEA are covered with adequate measures, e.g., adequacy decisions, Standard Contractual Clauses, etc.

⁶⁶ Ibid, Art. 32.

⁶⁷ Ibid, Art. 5(1)(e).

⁶⁸ Ibid, Art. 28.

3.3.15 Anonymisation and pseudonymisation

Anonymisation and pseudonymisation are security measures which may be used by the members of the Consortium in order to comply with the data security principle pursuant to Article 32 of the GDPR.⁶⁹ It is, however, important to establish a distinction between the two security measures in order to clarify when each of the measures may be relevant to use throughout the lifecycle of the project. The choice in using the two techniques is dependent on the specific situation and the level of data privacy protection which is required. Article 32(1)(a) of the GDPR recognises pseudonymisation as a way of achieving the security principle.⁷⁰ As previously mentioned, data is considered to be anonymised only when it is not possible to achieve identifiability of the person to whom the data belongs.⁷¹ Pseudonymisation instead entails a process which replaces identifiable information with a key that can be linked back to the original person with extra information. Essentially, this means that the pseudonymisation of data can enable data to be identifiable where more information is provided, however, anonymisation prevents the re-identification of data.⁷² The members of the Consortium will carefully evaluate the differences and feasibility of using such security measures.

3.4 Draft Artificial Intelligence Act

KT4D is a project about AI, but also one that proposes to use AI for some of its tasks. For that reason, we also give consideration of the potential impact on our data management that may stem from the introduction of the proposed Artificial Intelligence Act (AI Act). This Act focuses on strengthening the rules around data quality, transparency, human oversight as well as accountability. It also seeks to address the ethical questions and the challenges of implementation in the various sectors where Artificial Intelligence may be applicable. The Act lays down the harmonised rules for the use and placing of AI in the market, and it prohibits certain types of AI systems which are considered to be particularly risky, also laying down the specific requirements for high-risk AI systems in addition to placing obligations for operators of such systems.⁷³ The Act applies to several persons including the users of AI systems who are physically present or established in the Union and the product manufacturers who put into service an AI system.⁷⁴

At the time of writing, the draft AI Act is in the triilogue phase, the last stage of the legislative process.⁷⁵ This section of the DMP will therefore be updated according to relevant progress made on the Act in the coming months.

⁶⁹ Ibid, 32.

⁷⁰ Ibid, Art. 32(1)(a).

⁷¹ L Feiler, N Forgo, M Nebel *Article 4(5). Pseudonymisation*, in *The EU General Data Protection Regulation (GDPR): A Commentary* (Christopher Kuner and others (eds), online edn, Oxford Academic, 2020).

⁷² Regulation (EU) 2016/679, Recital 26.

⁷³ Artificial Intelligence Act, Art. 1.

⁷⁴ Ibid, Art 2.

⁷⁵ Luca Bertuzzi 'EU Council sets path for innovation measures in AI Act's negotiations', (*EurActiv*, 10 July 2023) <https://www.euractiv.com/section/artificial-intelligence/news/eu-council-sets-path-for-innovation-measures-in-ai-acts-negotiations/> accessed on 10 July 2023.

3.4.1 General classifications of AI systems

The AI Act outlines rules which follow a risk-based approach.⁷⁶ It suggests that a lighter legal regime applies to AI applications with a negligible risk, and that applications with an unacceptable risk are banned. Unacceptable risks are those that are considered to be a clear threat to the safety, livelihoods and rights of the people. They include AI systems or applications that manipulate human behaviour to circumvent users' free will (e.g., toys using voice assistance encouraging dangerous behaviour of minors) and systems that allow 'social scoring' by governments.⁷⁷ Some examples of high-risk AI systems listed in Annex III of the AI Act include: the use of AI in employment, workers management and access to self-employment (e.g., CV sorting software for recruitment procedures); essential private and public services (e.g., credit scoring denying citizens opportunity to obtain a loan); and law enforcement that may interfere with people's fundamental rights (e.g., evaluation of the reliability of evidence).

The High-Risk AI system must undergo an approved conformity assessment and continuously comply with AI requirements as set forth in the Act during its lifecycle. For certain systems an external notified body will be involved in the conformity assessment audit. Registration of the high-risk AI system is expected to take place in a dedicated EU database.⁷⁸ Limited risks AI systems have specific transparency obligations: when using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back. Lastly, the vast majority of AI systems fall into the category of minimal risk. The draft Regulation does not intervene here, as these AI systems represent only minimal or no risk for citizens' rights or safety.

3.4.2 Obligations on the providers of high risk systems

The draft AI Act places obligations on the providers of high risk systems which include: ensuring that the high-risk AI systems comply with the requirements of high-risk systems; indicating their name, registered trade name or registered trademark, the address at which they can be contacted on the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation; having a quality management system in place; and keeping the documentation as required in the Act.⁷⁹

3.4.3 Requirements of training machine learning models under the Act

This section provides a list of the requirements of training machine learning models under the Act and highlights the requirements that will be relevant in developing the AI system in the project.

3.4.4 Prohibited AI practices

The Artificial Intelligence Act places a prohibition on some AI practices which include:

- Placing or putting on the market an AI system that deploys subliminal techniques beyond a person's consciousness with the objective to or the effect of materially distorting a person's behaviour in a

⁷⁶ European Commission 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (*European Commission press release*, 21 April 2021) https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682 accessed on 3 July 2023.

⁷⁷ Artificial Intelligence Act, Annex III (High Risk AI systems referred to in Art. 6(3)).

⁷⁸ Kop, Mauritz, 'EU Artificial Intelligence Act : The European Approach to AI' (*Transatlantic Antitrust and IPR Developments*, Stanford Law School) (2021) < <https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/>> accessed on 26 April 2023.

⁷⁹ Artificial Intelligence Act, Art. 16.

manner that causes or is reasonably likely to cause that person or another person physical or psychological harm;

- Placing or putting into the market an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, disability or a specific social or economic situation, with the objective or the effect of materially distorting the behaviour of a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm;
- Placing or putting on the market an AI system for the evaluation or classification of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to detrimental or unfavourable treatment of certain natural persons or groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected and/or that is unjustified or disproportionate to their social behaviour or its gravity; and
- The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces by law enforcement authorities or on their behalf for the purpose of law enforcement unless it is necessary in achieving the following objectives: the targeted search for specific potential victims of crime; the prevention of a specific and substantial threat to the critical infrastructure, life, health or physical safety of natural persons or the prevention of terrorist attacks; and the localisation or identification of a natural person for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences.⁸⁰

3.4.5 Specific requirements for AI systems

The Artificial Intelligence Act outlines specific requirements to be implemented in AI systems which are illustrated below.

3.4.5.1 Establishment of a risk management system

A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems. The risk management system shall be understood as a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service.⁸¹

3.4.5.2 Data and data governance

Article 10 of the Act outlines that high-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets. Training, validation and testing data sets shall be complete, relevant, representative, and to the best extent possible, free of errors and shall take into account to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used.⁸²

⁸⁰ Ibid, Art. 5

⁸¹ Ibid, Art. 9.

⁸² Ibid, Art. 10.

3.4.5.3 A record of technical documentation

Article 4(b) of the draft Act places a requirement on the providers of the AI systems to keep a record of technical documentation at the disposal of the national competent authorities for a period ending 10 years after the general purpose of the AI system is placed on the Union market or put in the service of the Union.⁸³

The technical documentation subject to this requirement are outlined in Annex IV of the Act and include:

- A general description of the AI system;
- A detailed description of the elements of the AI system and of the process for its development;
- Detailed information about the monitoring, functioning and control of the AI system; a detailed description of the risk management system in accordance with Article 9 (Risk management system);
- A description of relevant changes made by the provider to the system through its lifecycle;
- A list of the harmonised standards applied in full or in part the references of which have been published in the Official Journal of the European Union;
- A copy of the EU declaration of conformity; and
- A detailed description of the system in place to evaluate the AI system performance in the post-market phase in accordance with Article 61 (post-market monitoring by providers and post-market monitoring plan for high-risk AI systems), including the post-market monitoring plan referred to in Article 61(3).⁸⁴

The Act, however, places an exemption on the provider who has explicitly excluded all high-risk (see above the qualification of AI systems as high risks) uses in the instructions of use or information accompanying the general purpose of the AI system.⁸⁵

3.4.5.4 Record keeping

Article 12 of the Act provides for the automatic recording of events ('logs') over the duration of the life cycle of the system for high-risk AI systems. Some of the information that shall be provided include the input data for which the search has led to a match, the reference database against which input data has been checked by the system and a recording of the period of each use of the system (start date and time and end date and time of each use).⁸⁶

3.4.5.5 Transparency and provision of information to users

The high-risk design systems shall be designed in a way that will ensure that their operation is sufficiently transparent with a view to achieving compliance with the relevant obligations of the user and of the provider. More importantly, Article 13 of the Act outlines that the high-risk AI system shall be accompanied by instructions for use in an appropriate digital format or otherwise will include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users. Some of the specified information include the identity and the contact details of the provider and, where applicable, of its authorised representative and the characteristics, capabilities and limitations of performance of the high-risk AI

⁸³ Artificial Intelligence Act, Art. 4(b).

⁸⁴ Ibid, Annex IV (Technical documentation referred to in Art. 11(1) of the AI Act).

⁸⁵ Ibid, Art. 4(c).

⁸⁶ Ibid, Art. 12.

system.⁸⁷ The development of the AI system by the members of the Consortium will also take into consideration the GDPR requirements on the transparency principle as described above.⁸⁸

3.4.5.6 Human oversight

Article 14 of the Act outlines that the high-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.⁸⁹

3.4.5.7 Accuracy, robustness, and cybersecurity

The draft AI Act requires that high-risk AI systems are designed and developed in such a way that they achieve, in light of their intended purpose, an appropriate level of accuracy, robustness and cybersecurity, and perform consistently in those respects throughout their lifecycle. This means that the AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems.⁹⁰ In the development of the AI system in the fulfilment of the project's objectives, the members of the KT4D Consortium will develop an appropriate level of accuracy and robustness throughout the lifecycle of the project. This will also take into consideration the GDPR requirements on the principle of accuracy which require that personal data is kept accurate, and where necessary up to date.⁹¹

3.4.5.8 Quality management system

The Act provides that the providers of high-risk AI systems shall put a quality management system which shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions.⁹²

3.5 ePrivacy Directive

The ePrivacy Directive,⁹³ a *lex specialis* to the GDPR, sets the standard for all entities which 'store or access information stored in the terminal equipment of a subscriber or user in the European Economic Area (EEA)'.

The ePrivacy Directive defines 'terminal equipment' as that which is directly or indirectly connected to the interface of a public telecommunications network to send, process or receive information, or a satellite earth equipment. While for the most part, the ePrivacy Directive applies to providers of publicly available electronic communication services and networks, Article 5(3) of the ePrivacy Directive establishes that 'the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia,

⁸⁷ Ibid, Art. 13.

⁸⁸ Regulation (EU) 2016/679, Art. 5(1)(a).

⁸⁹ Ibid, Art. 14.

⁹⁰ Kop, Mauritz, 'EU Artificial Intelligence Act : The European Approach to AI' (*Transatlantic Antitrust and IPR Developments*, Stanford Law School) (2021) < <https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/>> accessed on 3 July 2023.

⁹¹ Regulation (EU) 2016/679, 5(1)(d).

⁹² Artificial Intelligence Act, Art. 17.

⁹³ Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002].

about the purposes of the processing.’ This framework applies in relation to the KT4D project in the deposit of cookies on the user’s terminal in accessing the KT4D website by users. To this end, a cookie policy has been developed by one of the project’s partners which will be reviewed by ICTLC to ensure compliance with the requirements of the ePrivacy Directive and the transparency-related aspects necessary to ensure that valid consent is acquired from visitors of the KT4D website.

4 FAIR DATA

This section mainly describes the provisions made or envisioned to accommodate the FAIR principles and thereby future usage of KT4D data resources. The FAIR principles of findability, accessibility, interoperability and reusability provide guidelines for good data management practices and prescribe how FAIRness in data is accomplished through technical research.

4.1 Findable

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

4.1.1 Data sources the project builds on

The diversity of tasks described above indicate an underlying diversity of data sources being accessed. In general, however, the project builds upon pre-existing published research and policy literature, new datasets designed and collected by the project team (in particular via human subject research), and data that will be used to train and test the AI system. At this point in the project, the exact nature of the data to be accessed in the latter of these two categories is still being determined and a full description will be available by the time of the first update of this plan.

4.1.2 Data repositories, hosting services

For most project purposes, we will use a Google Drive (hosted via the ADAPT Centre, and therefore GDPR compliant) for internal sharing of data, Zenodo for the open sharing of reports and deliverables, and Zotero for the internal (and potentially, eventually public) sharing of bibliography and reference material. The project website is hosted and managed by project Partner 2, Trust IT. Where more specialised requirements

arise in the course of the development of the project (e.g. GitLab, hosting for ML activities, etc.) this will be described in the first update of this plan.

4.1.3 File naming conventions

At this point in the project lifecycle, there is no strict naming policy in place. At the level of work packages and within work packages, task leaders will decide on the optimum level of granularity in this respect and will ensure creators and curators harmonise their files along a coherent, shared policy. Relevant guidelines will be shared and coordinated across WPs.

4.2 Accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

How will the data be made accessible (e.g. by deposition in a repository)? What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)? How will the identity of the person accessing the data be ascertained?

4.2.1 Open Access, open data policy

Papers and reports resulting from the KT4D project will be published Open Access and a copy of them will be deposited in the Zenodo collection of the project. Data produced will also be made available Open Access unless legal or ethical restrictions make this impossible.

4.2.2 Documentation

All data will be accompanied by a codebook or datasheet to make the conditions of its collection and reuse absolutely clear. More complex objects comprising software and data (such as the Digital Democracy Lab) will be fully documented technologically, and will also be accompanied by a handbook describing its use and limitations.

4.3 Interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e., adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

4.3.1 Technical and social interoperability

Where possible, research teams will use standardised descriptors, knowledge organisation frameworks and vocabularies to structure and describe any newly created data, so as to enhance the potential for its later federation or reuse. We will also avoid the use of any proprietary data formats that do not allow export or reversal (e.g., the native formats of some Qualitative Data Analysis packages).

4.3.2 Legal interoperability

Ethical and legal issues associated with the KT4D project are discussed in more detail in the Project Deliverables D8.1 (H - Requirement No. 1), D8.2 (POPD - Requirement No. 2), D8.3 (AI - Requirement No. 3) and D8.4 (OEI - Requirement No. 4) and in Section 2, Legal Framework, of this document.

4.4 Reusable

How will the data be licensed to permit the widest re-use possible?

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

How long is it intended that the data remains re-usable? Are data quality assurance processes described?

4.4.1 Upstream reuse - obtaining reuse rights

In most cases, the reuse conditions will be clear to the research team from the point of access, in particular as pertains to published research (for which we would only be resharing metadata at most, which is generally possible under a CC-0 licence), and training data for the Digital Democracy Lab system. In this latter case, particular attention will be paid to the legal conditions under which the data was gathered, given the likely sensitive nature of such data.

4.4.2 Downstream reuse

The KT4D project is committed to the principle of open data, and to using the open platforms we have established to the maximum allowable extent. We must recognise, however, that some of the datasets we create will contain personal data from research participants, and this might restrict our ability to make data open. Further information about the project level consent procedures is available in project Deliverable 8.1.

4.4.3 Clear licensing policy

The KT4D project is committed to make available datasets as project outputs under a CC-BY 4.0 licence by default. This licensing policy will be implemented in a way that is easily understandable both for humans and for machines.

4.4.4 Sustainability plan for the project outputs

Sustainability of the project outputs will be ensured by the coordinating institution, which will ensure project outputs are available on reliable, durable open repositories to the greatest extent possible.

5 Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

5.1 DMP development timeline and responsibilities to update it

The KT4D project aims to use this DMP as a project management tool that facilitates developing a common understanding and shared data management solutions across the project participants. As such, all WPs of the project contribute to it via meetings with the WP leaders and written consultations carried out in an iterative fashion. T 1.1 (led by TCD) will be in charge of coordinating and leading regular updates of this DMP. The DMP will be a living document throughout the duration of the KT4D project with the first version at M6 and subsequent updates at M18 (version 2) and M36 (version 3).

5.2 Resources allocated in project budget

The costs for making the data associated with the KT4D project FAIR are covered by the grant. A total of 6 PMs is reserved for dealing with all issues around data management. During the project lifetime, TCD will guide the process that ultimately leads to the deposit of relevant components of the project in data repositories.

6 Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

Secure, GDPR compliant, cloud storage options are provided at the project level for various data types, as described above. Task leaders are expected to have additional back-up and storage policies in place for any local data following their own, local policies and back-up protocols using institutional cloud storage solutions (handled by their IT departments) and following the rules of GDPR. In case a partner institution does not have sufficient infrastructural components in place for secure storage, it is possible to coordinate with another project partner.

As discussed above under '4.1 Findable', data and software outputs will be deposited and made openly available on the long term in trusted repositories, to the greatest extent possible. More information about any barriers to this aim will be captured in this document as it is developed.

7 Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

The Ethical aspects of the KT4D project are described in detail in the Deliverables 8.1-8.4, and therefore not repeated here.

8 Conclusion and future work

Data management is not a task that can be completed, or indeed even fully imagined, at the start of a project. For that reason, the project team will consistently monitor its realisation of the ambitions expressed here, and its ever-developing understanding of the data requirements implied by these ambitions. The relevant data protection framework has been taken into consideration in the framework to ensure that data will be processed lawfully and according to the highest standards throughout the project's lifecycle. This first version of the DMP has also described the data management life cycle for all datasets to be collected, processed and/or generated by a research project, following the EU's guidelines (e.g., FAIR principles). This deliverable will elaborate the processing, storage, and the possible reusing of the data sets within the project. The next versions of the DMP will present an updated picture of the implementation of the DMP within the project and the related efforts. As a living document, the Consortium will make further updates to the DMP in light of the consortium's objectives. Essentially, the DMP will serve to fulfil the accountability principle pursuant to Article 5(2) of the GDPR.

9 References

No	Description/Link
R1	European Data Protection Supervisor, AEPD-EDPS joint paper on 10 misunderstandings related to anonymisation, 27 April 2021.
R2	Luca Bertuzzi 'EU Council sets path for innovation measures in AI Act's negotiations', (<i>EurActiv</i> , 10 July 2023) https://www.euractiv.com/section/artificial-intelligence/news/eu-council-sets-path-for-innovation-measures-in-ai-acts-negotiations/
R3	Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002].
R4	European Commission 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (European Commission press release, 21 April 2021) https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682
R5	European Commission Directorate-General for Research & Innovation, 'H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020' Version 3.0 of 26 July 2016 https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
R6	European Data Protection Supervisor, 'A Preliminary Opinion on data protection and scientific research', (6 January 2020).
R7	Information Commissioner's Office, 'Principle (a): Lawfulness, fairness and transparency', https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-protection-principles/a-guide-to-the-data-protection-principles/the-principles/lawfulness-fairness-and-transparency/#:~:text=At%20a%20glance&text=You%20must%20use%20personal%20data,will%20use%20their%20personal%20data
R8	Kop, Mauritz, 'EU Artificial Intelligence Act : The European Approach to AI' (<i>Transatlantic Antitrust and IPR Developments</i> , Stanford Law School) (2021) < https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/
R9	Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence) and Amending Certain Union Legislative Acts
R10	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1.
R11	Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)
R12	van Geuns, Jonathan and Ana Brandusescu, 'Shifting Power Through Data Governance' (Mozilla Insights, September 2020) < https://openfuture.eu/wp-content/uploads/2021/06/shifting-power-mozilla.pdf >

Appendix I. KT4D Record of Processing Activities v1.0



KT4D DATA PROCESSING ACTIVITIES

Introduction

This document has the purpose of mapping the processing activities carried out by partners of the Consortium in the context of KT4D project. Given that this document contains the mandatory elements of Art. 30 GDPR, we will use the information contained therein to perform Risk Assessments on data processing activities and Data Protection Impact Assessments (DPIAs) where necessary. When processing personal data please take into consideration the following principles:

- **Minimization:** You must ensure the personal data you are processing is adequate, i.e., sufficient to properly fulfil your stated purpose, relevant, i.e., it must have a rational link to that purpose, and limited to what is necessary, i.e., you do not hold more than you need for that purpose. Essentially, this means that you should not gather any information that you do not strictly need for the purposes of the project and the purposes should be listed clearly in relation to the data that is collected. This could also mean that when you can accomplish the aims of the research/project without using personal data, i.e., you could use anonymized data, that you should.
- **Data security:** Data should be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures ("integrity and confidentiality").
- **Storage limitation:** You must not keep personal data for longer than you need it; you need to think about - and be able to justify - how long you keep the personal data for which will be dependent on the purposes of keeping the data; you need a policy setting standard retention periods wherever possible, to comply with documentation requirements; you should also periodically review the data you hold, and erase or anonymize it when you no longer need it; you must carefully consider any challenges to your retention of data and remember that individuals have a right to erasure if you no longer need the data. It is, however, possible to keep personal data for longer if you are only keeping it for public interest archiving, scientific or historical research, or statistical purposes. We can help you with this evaluation.
- **Accountability:** The principle of accountability requires you to take responsibility for what you do with personal data and in terms of your compliance with data protection law. You should implement appropriate measures and keep good records to be able to demonstrate your compliance.
- **Validity of consent:** Consent must be demonstrable and linkable to the person who expressed it (it's necessary to keep a record of the consent), and such consent must be informed (data subjects/research participants must be informed in a clear way about how their data will be used and they must understand the consequences of providing their consent).
- **Avoid transferring personal data outside EU:** Unless strictly necessary, we must avoid transferring personal data to the USA, this means, e.g., that we should avoid storing personal data of people interviewed on Google or any other storage/tool provided by a US-based provider.

In case of any doubts concerning the lawfulness of personal data processing you are planning, please contact the project leader and ICTLC at: kt4d@ictlc.com

Table 7 – Partner personal data mapping instructions

KT4D Partner name	Enter your organization's name
Activity Name	Provide an identifier for the activity (e.g., Task T1.4)
Activity Description	Describe the activity carried out with personal data (e.g., interviews, focus groups etc.)
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	i.e., we will reach out to data subjects to request they participate in the research (contacting via email), we then need to interview them to understand...
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	i.e., we will process name, surname, age, voting patterns, social media and internet usage information, political opinions, photos, etc.
What type of data subjects will be involved?	i.e., students, clients of your organization, colleagues, academics, etc.
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	Note that the data of vulnerable individuals which include minors, the mentally ill, immigrants, the elderly, etc. require special protection.
From where will personal data be collected?	Describe how the data gathering/collection activity will take place. i.e., directly from the data subject or from other parties?
What is the envisioned data flow?	Please provide details on the envisioned data flow, i.e. researchers will reach out to the data subjects who will respond via email, then they will meet on location and we will have specific discussions which will be recorded by the researchers and which will be analyzed using X software, annotated by Y, and uploaded to the KT4D repository, etc.
How will data subjects be informed about the data processing activities? (i.e., information notice).	Please provide information, i.e., on informed consent.
Do you have an information notice to provide to data subjects?	Please provide us with a copy of the information notice used.
What are the purposes of processing (i.e. why do you need the data you are collecting)?	Be specific about each piece of data, i.e., we need their name to prove consent has been provided, we need their age because..., we need to know their education level because..., we need to know if they vote because..., etc.)
What is the legal basis for processing?	(The GDPR provides for six legal bases which can be relied upon to process personal data (see Art. 6 GDPR), these include consent, for the performance of a contract, to comply with a legal obligation, for the purposes of legitimate interests of the controller, etc.)
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent from presented to the data subjects.	Answer Yes or No and provide information on the consent form presented to data subjects. NB: You may need to collect informed consent for research purposes
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	Note that then entity that decides how data are processed is the data controller.
How long will data be stored?	Please also consider compliance with the FAIR data principles.
Where will data be stored?	i.e. on a shared drive, on paper-based materials.
What organizational and technical security measures are applied to protect the personal data?	i.e. the data repository is protected by MFA, access is restricted to those who strictly require access to the data, data is encrypted, pseudonymized, stored in a file cabinet which is locked in a locked room, etc.
Will personal data be shared with other partners and or other third parties?	Answer Yes or No and provide information on who the data will be shared with if data will be shared.
Will personal data be transferred outside EU? If yes, please specify where.	Note that if you use US-based providers such as Google, personal data will inevitably be transferred to the United States.
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	
Other relevant information	

Table 8 - KT4D partner personal data form

KT4D Partner name	Enter your organization's name
Activity Name	Provide an identifier for the activity (e.g., Task T1.4)
Activity Description	Describe the activity carried out with personal data (e.g., interviews, focus groups etc.)
What non-personal research data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	I.e., we will reach out to data subjects to request they participate in the research (contacting via email), we then need to interview them to understand...
What research data will you collect or produce?	I.e., we will process name, surname, age, voting patterns, social media and internet usage information, political opinions, photos, etc.
Where will the data be stored? How will it be accessed/secured? What measures do you have in place to ensure against data loss	eg. institutional cloud service, personal hard drive, project level services
What data formats will you use? How will the data be structured/organised (metadata, filename convention, versioning, application of any relevant standards)? Roughly how much data do you expect this to be?	eg., pdf, txt, mpg., etc.;
From where will this data be collected? Under what kind of reuse license?	Describe how the data gathering/collection activity will take place, ie. from institutional databases; license may be CC-0, CC-by or other
What if any additional measures will you take to enhance the FAIRness of the data?	eg. deposit alongside contextual material
Will this/can this data be made available open access?	
Other relevant information	

Table 9 – KT4D Partner Research Data Form

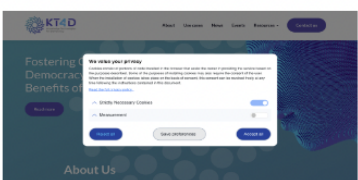
KT4D Partner name	Beyond the Horizon (BtH)
Activity Name	Work Package 2
Activity Description	Task 2.3 Events management, including Use Case consultation and co-creation meetings
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	Event registration requires to process personal data, such as participant name, surname and email.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	Apart from mandatory fields in any event registration, i.e. name, surname and email address, optional fields can be added for research and statistical reasons in consultation with the partners, such as age (range), job affiliation or job role. We will not process any special category of data.
What type of data subjects will be involved?	Academics, researchers, general public, policy makers.
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	Potentially minors and elderly.
From where will personal data be collected?	Via a form on the website.
What is the envisioned data flow?	The data, once submitted through the webform will be safely stored in the internal database.
How will data subjects be informed about the data processing activities? (i.e., information notice)	Privacy Policy on the website (https://kt4democracy.eu/privacy-policy-full).
Do you have an information notice to provide to data subjects?	Privacy Policy on the website (https://kt4democracy.eu/privacy-policy-full).
What are the purposes of processing (i.e. why do you need the data you are collecting)?	We need them in our Dissemination, Exploitation and Communication activities.
What is the legal basis for processing?	The legal basis for processing refers to Art.6 comma a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes (GDPR).
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent form presented to the data subjects	The consent is collected via a Privacy Form as one visits the project website. 
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	TRUST-IT SRL is the KT4D Data Controller.
How long will data be stored?	Data will be stored during the project lifecycle and five more years upon project conclusion, as requested from the European Commission.
Where will data be stored?	Data will be stored in a cloud platform, provided by AWS, located in Europe.
What organizational and technical security measures are applied to protect the personal data	Organizational Security Measures: Access Control Policies: Strict access control policies are enforced to ensure that only authorized personnel can access the personal data located in our Cloud provider. This includes the use of strong authentication mechanisms, such as multi-factor authentication (MFA), and role-based access control to limit privileges. Data Privacy Policies: Organizations establish comprehensive data privacy policies that outline the handling, storage, and processing of personal data. These policies are designed to comply with relevant data protection regulations, such as the General Data Protection Regulation (GDPR) Technical Security Measures: Encryption: Personal data is encrypted both in transit and at rest. Transport Layer Security (TLS) or Secure Sockets Layer (SSL) protocols are used to encrypt data during transmission, while data at rest is typically encrypted using technologies like Amazon S3 server-side encryption or AWS Key Management Service (KMS). Network Security: AWS implements a robust network security infrastructure to protect personal data. This includes network segmentation, firewalls, intrusion detection and prevention systems (IDS/IPS), and distributed denial-of-service (DDoS) protection mechanisms. Data Backup and Disaster Recovery: Regular data backups are performed to ensure data availability and integrity. AWS offers various backup and disaster recovery services, such as Amazon S3, and AWS Backup, to facilitate reliable data protection and recovery processes. Logging and Monitoring: AWS provides extensive logging and monitoring capabilities to track and detect any suspicious activities or unauthorized access attempts. Services like AWS CloudWatch and AWS GuardDuty enable real-time monitoring, log analysis, and alerting. Physical Security: AWS datacenters are equipped with stringent physical security measures, including access controls, surveillance systems, and 24/7 on-site security personnel. These measures ensure that only authorized individuals can physically access the datacenters.
Will personal data be shared with other partners and/or other third parties?	Data will be shared only with Partners and third parties listed as Data Processor in the Privacy Policy Statements.
Will personal data be transferred outside EU? If yes, please specify where.	Data are stored and managed in the EU.
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	No

Table 10 – Beyond the Horizon (BtH) data processing activities

KT4D Partner name	Fundacion Cibervoluntarios
Activity Name	Task T6.3 - T6.1
Activity Description	Co-creation workshops, group interviews, tool testing and validation
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	I.e., we will reach out to data subjects (or organizations that to request they participate in the research) contacting them via email, phone call and through social media and newsletter channels. We then will need them to participate in two sessions over all the project which will include a workshop with group interviews and a validation UX session.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	age, gender, and opinions towards political participation
What type of data subjects will be involved?	Adults, cybervolunteers, minors (16-17) seniors and professionals
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	There might be some migrants in the sample, and minors over 16, as well as elderly people but without any cognitive impairment.
From where will personal data be collected?	data gathering/collection activity will take place directly from the data subjects and interactions in the form of notes.
What is the envisioned data flow?	researchers will reach out to the data subjects who will respond via email, then they will meet on location and we will have specific discussions which will be annotated by the researchers and which will be analyzed without any specific software aside from Office text processor (since it will be qualitative data)
How will data subjects be informed about the data processing activities? (i.e., information notice)	Participants will have an adapted information sheet at least 10 days in advance and a signed consent with the same notice. Before the session start, the researcher will make sure the participants have read and understand the information in the information sheet and signed the consent knowing its content and implication.
Do you have an information notice to provide to data subjects?	https://docs.google.com/document/d/1Yf82UihU78XOorLEzrv1DVTbedE/edit
What are the purposes of processing (i.e. why do you need the data you are collecting)?	We will collect age (due to its correlation with differentiated use of digital tool), habits of political participation and use of digital platform/tools as well as details of their interaction and opinions about the education materials and games of KT4D
What is the legal basis for processing?	The legal basis is consent, one of the 6 GDPR legal basis
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent form presented to the data subjects	https://docs.google.com/document/d/1RA33ssfOWTxUqXrUEARgVoZB4SSDEbjQ/edit
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	For CIB work, CIB decides how data is stored in compliance with European and national law. Cibervoluntarios DPO is Jorge Rastrilla Caballero jorge.rastrilla@cibervoluntarios.org
How long will data be stored?	Data will be stored for the shortest time possible after the project's end
Where will data be stored?	Data will be stored paper-based in CIB's offices and scanned in digital format in protected servers
What organizational and technical security measures are applied to protect the personal data	I.e. the data repository is protected by MFA, access is restricted to those who strictly require access to the data, data is encrypted, pseudonymized, stored in a file cabinet which is locked in a locked room, etc.
Will personal data be shared with other partners and/or other third parties?	Data about political participation, age and use of digital technologies may be shared with other partners of the consortium only for research purposes and only with a legal declaration that this data will not be shared to third parties. No data will be shared to third parties.
Will personal data be transferred outside EU? If yes, please specify where.	No
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	It will not
Other relevant information	

Table 11 – Fundacion Cibervoluntarios data processing activities

KT4D Partner name	Demos Research Institute oy
Activity Name	Use case 3 / T4.3 & T5.1
Activity Description	Use case workshops & Delphi study
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	Participants i.e. policymakers and policy experts will be mapped and identified based on existing AI governance networks and policy analysis of openly available online documents, including white papers, legislation and regulation. Policymakers identified will be invited to take part in the use case and a delphi study through email invitations based on their organisational affiliation. After the first workshop the participant's emails (including all individuals who signed up) will be used as contact information to send out a delphi survey form. The participants will be asked their informed consent about the use of their personal data in processing the data collected via the form. The individual respondents cannot be identified based on the reporting of results. There will be 3 workshops in total with similar procedure for data processing.
What personal data will you	We will process name, surname, organisation, professional title and email. No special categories of data will be processed.
What type of data subjects will be	Policymakers and experts from EU institutions, public organisations and civil society
Will you process the personal	No
From where will personal data be	For the initial mapping of potential participants, we will collect publicly available email addresses from organisation websites. Personal data will be collected directly from the data subject once they sign up for the workshop.
What is the envisioned data flow?	Researchers will contact the data subjects who will respond via email. Researchers and data subjects will meet in an on-site workshop in Brussels. The discussions from the workshops will be documented by taking notes of the content of the discussion without an explicit identification of individual speakers. The notes will be stored by the Demos team in GDPR compliant and secure location based on the Demos data storage practises.
How will data subjects be	On informed consent collected in connection to signing up in the workshop.
Do you have an information	the Demos privacy policy https://demoselsinki.fi/privacy-policy/
What are the purposes of processing (i.e. why do you need the data you are collecting)?	We will need their name to prove consent has been provided, their organisation to ensure key parties have been included in the discussion, and titles to specify which kind of roles, seniority level and perspectives the participants represent.
What is the legal basis for	Consent
If the legal basis relied upon is	https://docs.google.com/document/d/1NUjcb5ms1tHLD0idj6AbnER5jijt0d/edit
Who/what entity decides how	Demos Research Institute oy
How long will data be stored?	The data will be deleted as soon as possible after the project's ending date.
Where will data be stored?	On a shared drive by the Demos team in GDPR compliant and secure location based on the Demos data storage practises.
What organisational and	the access is restricted to those who strictly require access to the data. Further, see our privacy policy point 11 - Usernames and passwords: Accessing the data in the register requires user-specific usernames and passwords
Will personal data be shared with	The list of participants with email addresses will be shared with Beyond Horizon for workshop organising purposes
Will personal data be transferred	Yes, based on the use of Google drive
Will your data processing activity	No
Other relevant information	N/A

Table 12 – Demos Research Institute data processing activities

KT4D Partner name	Demos Research Institute oy
Activity Name	T4.1, T4.3, T5.1
Activity Description	Literature review and document analysis in each task. We will compose a corpus on ethics and AI from the perspectives of e.g. global and regional governance, EU digital policy, self-assessment guidelines
What non-personal research data	Analysis of publicly available documents from open sources and an academic literature review
What research data will you collect or	We will collect and produce data on the existing literature, guidelines, debates and resources on AI ethics, AI and data governance (existing practices, critical questions, implications where possible), EU
Where will the data be stored? How will it	The lists of references created for the literature review and document analysis will be stored in the KT4D Zotero library. The governance framework v1.0 (WP5) will be also made available for the consortium.
What data formats will you use? How will	pdf, txt. Documents will be organised by file name and made searchable for the KT4D consortium members.
From where will this data be collected?	From institutional databases, online research repositories, websites
What if any additional measures will you	Zotero will be used for storage of the organised sources lists and possibly made public at the end of the project.
Will this/can this data be made available	A joint decision about this will be made by the consortium
Other relevant information	N/A

Table 13 – Demos Research Institute research data

KT4D Partner name	Democratic Society
Activity Name	WP7 (specifically task 7.2 and 7.3)
Activity Description	Digital Democracy Lab workshops, group interviews and ethnographic documentation and observation
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	We will engage with participants from different stakeholder groups during the Digital Democracy Lab workshops in form of participatory observation and group discussions. Observations and data collecting during the workshops will be attained without personal indicators such as names, age etc. Part of the Digital Democracy Lab will be built on the Democracy Lab Demonstrator system, where personal or invented data (personas) will be used to understand how data is collected, how it is processed and used. The criteria or using personal profiles during the workshop is currently being evaluated and formulated in Task 7.1 to ensure data protection and accountability.
What personal data will you process? Will you process special categories of personal data (e.g. ...)	For the workshops, group interviews, ethnographic documentation and observation, we will not collect personal data.
What type of data subjects will be involved?	The Digital Democratic Lab will work with different stakeholder groups during the use cases: policy makers, software developers, citizens and citizen facing civil society organizations.
Will you process the personal data of vulnerable people	
From where will personal data be collected?	In the Digital Democratic Demonstrator, we will use external/third party data for ML-training. What data and how data protection will be implemented will be discussed and implemented in Task 7.1/7.2 and the Data Processing plan will be updated.
What is the envisioned data flow?	Please provide details on the envisioned data flow, i.e. researchers will reach out to the data subjects who will respond via email, then they will meet on location and we will have specific discussions which will be recorded by the researchers and which will be analyzed using X software, annotated by Y, and uploaded to the KT4D repository, etc. The participation recruitment is not in the hands of Demosc. The data that will be collected during the Democracy Labs will be in form of observation journaling, possible collective discussions/interviews. This material will then be analyzed and annotated by using MaxQDA or AirTable and uploaded to the KT4D repository.
How will data subjects be informed about the data processing activities? (i.e., information notice)	Participants will be asked to sign only a Democratic Society's consent form or a collective KT4D used for all use cases. The participants will be asked to read and sign the consent form at the beginning of each Digital Democracy Lab. The consent form will highlight which data is being collected by the KT4D project, how their data is being used, and ask for consent to be observed, and photographed during the event itself. The consent forms will be translated into local languages by the partners on the ground, and the template document will be stored on a shared drive. Completed documents will be stored on the Demosc Google Drive, protected by a password to be added to Demosc's LastPass account. A consent form for guardians of underage participants will be shared with them during the sign-up process to ensure the safety and security of children and youth participating. The form will ensure that guardians agree to the participation of youth participants. If participants choose to not be photographed during the Digital Democracy Lab, they will receive a sticker on their name badge to identify them for the photographer. As for the Digital Democracy Demonstrator, the participants will receive an in-depth understanding of how their personal data is used (if it is used) and all personal data will be deleted at the end of the Workshop.
Do you have an	Please provide us with a copy of the information notice used.
What are the purposes of processing (i.e. why do you need the data you are collecting?)	Be specific about each piece of data, i.e., we need their name to prove consent has been provided, we need their age because..., we need to know their education level because..., we need to know if they vote because..., etc.)
What is the legal basis for processing?	Consent The consent forms will either be digital or printed e.g. at the event itself. The paper forms will be stored on only the Demosc's Google Drive account. The original version shall be destroyed once the scanned version is available, and only the KT4D team will have access to them.
If the legal basis relied upon is consent, did you collect consent? If so, also Who/what entity decides how personal data are processed? Please specify	https://docs.google.com/document/d/1RA33ssfOWTxUqXrUEARgVoZB4S5DEbJQ/edit Democratic Society
How long will data be stored?	Data will be stored only for the purpose of the KT4D project activities and deleted a maximum of 3 months after an event have taken place – with us retaining only anonymized data that is needed for reporting or project delivery.
Where will data be stored?	On a shared drive by the DemSoc team in GDPR compliant and secure location based on DemSoc's data storage practises. The consent forms will either be digital or printed e.g. at the event itself. The paper forms will be stored on only the DemSoc's Dropbox account. The original version shall be destroyed once the scanned version is available, and only the DemSoc team will have access to them. Data will be stored only for the purpose of the PaCE project activities and deleted a maximum of 3 months after an event have taken place – with us retaining only anonymized data that is needed for reporting or project delivery. The exemption is the data of individuals that want to participate in the European Lab.
What organizational and technical security	The access is restricted to those who strictly require access to the data. The data will be password protected.
Will personal data be shared with other partners and/or other	The anonymized data will be shared with Hybrid Core and Trinity Collages for the development of Task 7.3
Will personal data be transferred outside EU? If yes, please specify where.	No
Will your data processing activity involve AI or machine learning	For the Digital Democratic Demonstrator ML-techniques will be used. For models and data sets see outcome of Task 7.1 and 7.2 and data processing plan from Hybridcore
Other relevant information	N/A

Table 14 – Democratic Society data processing activities

KT4D Partner name	Hybrid Core
Activity Name	WP6 (Task 6.2), WP7 (Task 7.2 and Task 7.3)
Activity Description	Building and deploying technical components to support the Digital Democracy Lab Demonstrator system.
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	The ability to upload and store user-provided data is a crucial feature for any platform that aims to support ethical user profiling. By creating a data cooperative, the platform can ensure that users have control over their data and that it is used in a transparent and accountable way. Tagging of training data and comparison with existing testing data is an important aspect of participatory machine learning design. This feature will allow users to contribute to the development of machine learning models and improve their accuracy and fairness. The ability to view, apply, and propose profiling cases is a key feature for promoting ethical user profiling. By providing a sandbox environment, users can explore different profiling techniques and understand how they work without risking any negative consequences. The use of participants' pre-existing online footprint, individually provided data, and fake personas is a sensible approach to creating user profiles while respecting privacy and confidentiality. The use of big data techniques such as usage mining and clustering can help to improve the accuracy and relevance of these profiles.
What personal data will you process? Will you process special data?	The user-provided personal data will be processed.
What type of data subjects will be involved?	Citizens, civil societies and organizations, policymakers etc.
Will you process the personal data of vulnerable people (minors, immigrants or refugees)?	No
From where will personal data be collected?	Data cooperatives, 3rd party data sources in compliance with GDPR.
What is the envisioned data flow?	The data that will be collected during the Democracy Labs will be in form of observation journaling, possible collective discussions/interviews.
How will data subjects be informed about the data processing activities? (i.e., information notice)	Participants will only be required to sign a consent form from Democratic Society or a collective KT4D used for all use cases. At the beginning of each Digital Democracy Lab, participants will be required to explore and sign the consent form. The consent form will specify which data is being collected by the KT4D project, how their data is being utilized, and request permission to be observed and photographed during the event. Partners on the ground will translate consent forms into local languages, and the template document will be stored on a shared drive. Completed documents will be saved to the Demsoc Google Drive, protected by a password that will be added to the Demsoc account on LastPass. To ensure the safety and security of children and adolescents participating, a consent form will be distributed to their parents or custodians during the registration process. The form will assure that youth participants have parental permission to participate. If a participant does not wish to be photographed during the Digital Democracy Lab, a label will be placed on their name badge to alert the photographer. As with the Digital Democracy Demonstrator, the participants will receive an in-depth comprehension of how their personal data is used (if it is used), and at the conclusion of the Workshop, all personal data will be deleted.
Do you have an information notice to provide to data subjects?	NA
What are the purposes of processing (i.e. why do you need the data you are collecting)?	Hy8 will not be collecting any data, but will use the collected and tagged data in order to train the ML algorithm for Democracy Lab.
What is the legal basis for processing?	Consent forms will be filled in accordance with GDPR.
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent form presented to the data subjects	No
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	Democratic Society
How long will data be stored?	Data will be stored solely for the purposes of the KT4D project activities and will be purged no later than three months after an event, with the exception of anonymized data required for reporting or project delivery.
What organizational and technical security measures are applied to protect the personal data	On a shared drive by the DemSoc team in GDPR compliant and secure location based on DemSoc's data storage practises. The access is restricted to those who strictly require access to the data and will be stored in GDPR compliant cloud servers for ML training purposes only.
Will personal data be shared with other partners and or other third parties?	No
Will personal data be transferred outside EU? If yes, please specify where.	No
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	The platform would require machine learning algorithms to facilitate the tagging of training data and the comparison of newly tagged training data with existing testing data. These algorithms could include techniques such as classification, regression, and clustering.
Other relevant information	

Table 15 – Hybrid Core data processing activities

KT4D Partner name	Institute of Urban and Regional Development
Activity Name	Work Package 3: Advancing the state of the art: Civic Participation and Knowledge Technologies
Activity Description	Task 3.4.: Interaction design for Participatory Design Session, Use Cases 2 and 3. Focus group research, social games
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	3-step sessions: 1. warm up in groups of 15-16 people, carefully selected (age, gender and other relevant backgrounds), 2. Discussion/Focus group (groups of 4 people), 3. Validation/co-creation using games and educational material. We will ask participants to give us their informed consent about the use of their personal data in processing the data collected.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	We will process name, family name, age, gender, e-mail address and other relevant data if necessary to let the participants take part to focus and to exercises co-creating materials and games and, further, take part to it
What type of data subjects will be involved?	City users of different ages, ethnic minorities,
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)	We do not exclude to involve vulnerable individuals in our research activities, in order to take better into consideration their vulnerability, we would like to use co-creation as potential source of tailoring our standardised research procedures.
From where will personal data be collected?	Directly from the data subject while she/he comes to take part to the session
What is the envisioned data flow?	Researchers will contact the representative data subjects so to inform and collect their interest in research project. Researchers and chosen data subjects will meet in an on-site 3-step session in Warsaw. The discussions from the sessions will be documented by taking notes of the content of the discussion without identification of individual speakers. The notes will be stored by the IRMIR team in GDPR compliant and secure location based on the internal IRMIR data storage policy.
How will data subjects be informed about the data processing activities? (i.e., information notice)	Personal data subjects will be informed about the research project activities and types of data which are relevant through an information notice which is attached to an informed consent form
Do you have an information notice to provide to data subjects?	Privacy Policy on KT4D website (https://kt4democracy.eu/privacy-policy-full)
What are the purposes of processing (i.e. why do you need the data you are collecting)?	We need their name to prove consent has been provided but also to let participants to focus and games interact among them, we need their age to further process data analysis, we need to know their e-mail address to let them know about the results of our research, all the data we have to collect are necessary for dissemination, exploitation and communication purpose.
What is the legal basis for processing?	The legal basis for processing refers to GDPR, art.6 a) which states the data subject has given consent to the processing of his or her personal data for one or more specific purposes.
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent from presented to the data subjects	Yes. The consent is collected via a Privacy Form as one visits the project website, we are currently working on the consent form in Polish
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	TRUST-IT SRL is the KT4D Data Controller
How long will data be stored?	The data will be deleted as soon as possible depending on the project final date.
Where will data be stored?	On a shared drive by the IRMIR team in GDPR compliant and secure location
What organizational and technical security measures are applied to protect the personal data	Access to the data repository is restricted to researchers who strictly require access to the data, accessing the data in the register requires user-specific usernames and passwords which are granted only for use by specific predefined employees of the controller. The system is protected by firewalls and other technical means. Our website is protected by an SSL certificate, which ensures a safe and secure connection. between your browser and the server. Data are stored in a file cabinet which is locked in a locked room.
Will personal data be shared with other partners and or other third parties?	Data will be shared only with partners to the project and third parties listed as Data Processor in the Privacy Policy Statements.
Will personal data be transferred outside EU? If yes, please specify where.	Data will not be transferred outside EU.
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	We will not use AI nor machine learning techniques in data processing activity
Other relevant information	No

Table 16 – Institute of Urban and Regional Development data processing activities

KT4D Partner name	Institute of Urban and Regional Development
Activity Name	Work Package 3: Advancing the state of the art: Civic Participation and Knowledge Technologies
Activity Description	Task 3.4.: Interaction design for Participatory Design Session, Use Cases 2 and 3. Focus group research, social games
What non-personal research data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	We will reach out to data subjects to request they participate in the research, contacting them via email, we will analyse publicly available documents from open sources and an academic literature review
What research data will you collect or produce?	We will collect self-report data in Focus and games, calling for participants' ideas, beliefs, preferences, etc.
Where will the data be stored? How will it be accessed/secured? What measures do you have in place to ensure against data loss	On password protected hard drive and on institutional cloud drive.
What data formats will you use? How will the data be structured/organised (metadata, filename convention, versioning, application of any relevant standards)? Roughly how much data do you expect this to be?	pdf, doc, txt. Documents will be organised by file name
From where will this data be collected? Under what kind of reuse license?	From institutional databases, online research repositories, websites and, concerning Focus sessions, participants will be recruited through online platforms.
What if any additional measures will you take to enhance the FAIRness of the data?	Not at this moment
Will this/can this data be made available open access ?	only anonymized data
Other relevant information	No

Table 17 – Institute of Urban and Regional Development Research data

KT4D Partner name	Strane Innovation (STRANE)
Activity Name	T3.2 & T3.3
Activity Description	Online experimental studies
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	T3.2 & T3.3 will involve data collection (through online platform connecting researchers and participants for academic studies), analysis (using the open-source statistical software R), storage (on open-access platforms) and the publication of its results in academic journals.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	We will collect participants' Prolific IDs (or other similar research platforms), as well responses to generic demographic questions (age category, level of education, socio-professional category, political orientation).
What type of data subjects will be involved?	Participants registered in academic research platforms such as Prolific (https://www.prolific.co/)
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	No
From where will personal data be collected?	Participants will be recruited through online platforms that connects researchers with volunteers who wish to participate occasionally in paid scientific studies, such as Prolific (https://www.prolific.co.)
What is the envisioned data flow?	The data flow involves the following steps: 1. We create a study on the Prolific platform; 2. Participants are recruited from the Prolific user base; 3. Enrolled participants take part in the study, and the data generated is collected by the Prolific platform, which ensures its validity and reliability; 4. Once a participant completes a study, they receive compensation. Prolific handles the payment process, transferring the funds to the participant's Prolific account; 5. We access the collected data through the Prolific platform, by downloading it in a .csv file; 6. The results of the statistical analysis, together with the anonymised datasets are made available in an open-access repository.
How will data subjects be informed about the data processing activities? (i.e., information notice)	All participants will be presented an information notice prior to the start of the study, which will detail the features of the study, including the contact details of the researcher(s), the purpose of the research and what participation involves, how the data will be used and kept secure, how the participants' privacy will be protected, risks of taking part in the study, compensation for taking part, and the possibility to withdraw from the study.
Do you have an information notice to provide to data subjects?	Yes (see Information and consent notice attached)
What are the purposes of processing (i.e. why do you need the data you are collecting)?	Research purposes, including the investigation of the links between prior beliefs and attitudes toward AI and big data.
What is the legal basis for processing?	The legal basis for the processing is the explicit and informed consent of the participants who will take part in the studies.
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent from presented to the data subjects	Yes (see Information and consent notice attached)
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	The Data Protection Officer of the University Paris Cité (to which WP3 leader Tiffany Morisseau is affiliated) will ensure compliance with data protection regulations for the experimental work carried out in T3.2 and T3.3. Contact details: dpo@u-paris.fr.
How long will data be stored? Where will data be stored?	The dataset will be stored on restricted access cloud-based servers (owned by Strane Innovation). The dataset will be stored for up to five years.
What organizational and technical security measures are applied to protect the personal data	The platforms that will be used to recruit participants, together with the content of the questions asked in the studies will make it impossible to cross-check the data to trace the identity of the participants.
Will personal data be shared with other partners and/or other third parties?	Only fully anonymised (in such a way that individuals are no longer identifiable) datasets will be shared with other partners and made open-access.
Will personal data be transferred outside EU? If yes, please specify where.	No
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	No
Other relevant information	NA

Table 18 – Strane Innovation (STRANE) data processing activities

KT4D Partner name	Strane Innovation (STRANE)
Activity Name	T3.2 & T3.3
Activity Description	Online experimental studies
What non-personal research data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	T3.2 & T3.3 will involve data collection (through online platform connecting researchers and participants for academic studies), analysis (using the open-source statistical software R), storage (on open-access platforms) and the publication of its results in academic journals.
What research data will you collect or produce?	We will collect behavioural data (e.g., response time, accuracy rates, etc.) and self-report data (questions about participants' attitudes, beliefs, preferences, etc.).
Where will the data be stored? How will it be accessed/secured? What measures do you have in place to ensure against data loss	The anonymised datasets will be stored and be made open access on the Open Science Framework platform (https://osf.io/).
What data formats will you use? How will the data be structured/organised (metadata, filename convention, versioning, application of any relevant standards)? Roughly how much data do you expect this to be?	The data of the experiments will be collected via Qualtrics and Google Form and stored using .csv and .txt formats. Roughly, we plan to collect data from a maximum of 5,000 participants.
From where will this data be collected? Under what kind of reuse license?	Participants will be recruited through online platforms that connects researchers with volunteers who wish to participate occasionally in paid scientific studies, such as Prolific (https://www.prolific.co.) Only anonymised datasets are delivered by these platforms.
What if any additional measures will you take to enhance the FAIRness of the data?	NA
Will this/can this data be made available open access?	Yes
Other relevant information	NA

Table 19 – Strane Innovation (STRANE) research data

KT4D Partner name	Trinity College Dublin
Activity Name	T1.1
Activity Description	Project meetings and documentation: setting up mailing lists and scheduling virtual meetings (e.g., Zoom)
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	In order to co-ordinate the project, TCD will collect email addresses for each partner involved in the project so we can set up mailing lists, a shared internal project communication platform and schedule virtual meetings over the course of the project.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	We will process the name, surname, email and organisation for each member of the consortium partners who are involved with or working on the KT4D project
What type of data subjects will be involved?	Members of the KT4D consortium partners who are involved in the project
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	N/A
From where will personal data be collected?	The data will be collected from each consortium partner via a Contacts Spreadsheet that each partner will be asked to complete
What is the envisioned data flow?	Each member of the KT4D consortium partners will be asked to supply their email and highlight the mailing lists they would like to be included in. A Contacts Spreadsheet created for this purpose will be kept in the KT4D shared Google Drive to be accessed by members of the KT4D consortium
How will data subjects be informed about the data?	In setting up the Contacts Spreadsheet, members of the consortium will be informed that the Contacts Spreadsheet and the information contained therein will be available in the KT4D shared Google Drive to be accessed by members of the KT4D consortium.
Do you have an information notice to provide to data subjects?	Email sent to consortium members: If you can please fill in this (contacts) list, I'll update the mailing lists and Google Drive folders. Moving forward, can you please let me know if there is a change in your organisation so I can add members or remove anyone who is moving on from KT4D.
What are the purposes of processing (i.e. why do you need the data you are collecting)?	In order to maintain up to date mailing lists and provide the correct access to the KT4D Google Drive folders, it is necessary to collect the name, email and organisation that each member belongs to.
What is the legal basis for processing?	For the performance of the project that each member of the consortium belongs to
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent from presented to the data subjects	N/A
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	As coordinator of KT4D, Trinity College Dublin determines how the personal data (names, emails and organisations) are processed. The Project Manager (Eva Power eva.power@adaptcentre.ie) has created a contacts spreadsheet for each partner to complete. This contacts spreadsheet is made available in the KT4D Google Drive folder and access is granted to the members of the KT4D consortium partners.
How long will data be stored?	The contact details will be kept for one year after the duration of the project. If a member of a consortium partners leaves their organisation, their email will be removed from the applicable mailing lists and their access to KT4D's Google Drive will be revoked.
Where will data be stored?	On KT4D's shared Google Drive that is hosted by ADAPT in TCD
What organisational and technical security measures are applied to protect the personal data	The consortium member's names, email and organisations are in the Contacts spreadsheet that is located in KT4D's shared Google Drive which is hosted by ADAPT. Private passwords are used by each member to gain access to their Google Drive accounts.
Will personal data be shared with other partners and/or other third parties?	No
Will personal data be transferred outside EU? If yes, please specify where.	Concerning Google, the organisation controls the data region (Europe) and by using Europe for all ADAPT members. All mailing lists and KT4D Google Drive is through the ADAPT membership. With respect to Zoom, certain data may be transferred outside of the EEA using the SCCs and other lawful transfer mechanisms.
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	No
Other relevant information	NA

Table 20 – Trinity College Dublin data processing activities

KT4D Partner name	Trinity College Dublin (TCD)
Activity Name	Task 3.1; Task 5.2; Task 6.1; Task 7.1
Activity Description	Task 3.1 (literature review: building a corpus of accounts documenting both historical and imagined interactions with knowledge technologies and their effects on individual self-determination); Task 5.2 (review of existing tools to guide software developers toward human-centred design); Task 6.1 (literature review: building a corpus on critical digital literacy in the context of civic and democratic participation); Task 7.1 (literature review: building a corpus on Participatory Algorithmic Accountability within the digital humanities and cultural critiques of knowledge technologies).
What non-personal research data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	We will put together a number of corpora, each one dedicated to one of the topics connected with each task (see above). In order to do so, we will survey the existing academic literature and organise the sources into specific categories so as to make them easily searchable by the members of KT4D team.
What research data will you collect or produce?	We will collect research data about the long history of knowledge technologies (T3.1), human-centred and participatory design (T5.2), critical digital literacy (T6.1), Participatory Algorithmic Accountability (T7.1) from scholarly journals and publications, but also from magazines, conference proceedings, websites, online resources.
Where will the data be stored? How will it be accessed/secured? What measures do you have in place to ensure against data loss	The lists of references created for the literature reviews (Tasks 3.1, 6.1, 7.1) will be stored in the Zotero library set up for the project. Where possible (meaning when the source is in open access or freely available), a copy of the source will also be stored in the Zotero library. The same will happen for the list of existing tools to guide software developers (Task 5.2), which will be stored in the Zotero library for WPS.
What data formats will you use? How will the data be structured/organised (metadata, filename convention,	The data collected will be stored in .pdf and .docx formats (corpora), and in .xlsx for the list of sources (bibliographies). We expect to provide tags for each source listed in the Zotero library, so that the content can be mapped into different areas of interest. The expected size of the data is still unknown and it will depend on how many relevant sources we will find for each different task.
From where will this data be collected? Under what kind of reuse license?	In order to survey the field and collect the academic and non academic sources needed, we will take advantage of online research repositories, library repositories, but also digital libraries, websites, online newspapers, etc. In addition to printed material and resources made available by physical libraries and archives.
What if any additional measures will you take to enhance the FAIRness of the data?	We will use Zotero in order to store the list of sources (see above) as an open platform that we can make public if we choose to at the end of the project. This will make the corpora we create easy to be access and re-use by other researchers. In Zotero, we will use tags so as to make the corpora more easily searchable.
Will this/can this data be made available open access ?	It is most likely that the Zotero library containing all the references organised by area/topic and provided with the relevant tags will be freely available at the end of the project. At the moment, we are still discussing what exactly will be made openly available (if everything, or just a selection of the corpora created during the project).
Other relevant information	N/A

Table 21 – Trinity College Dublin research data

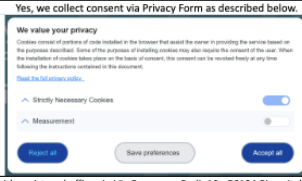
KT4D Partner name	Trust-IT
Activity Name	Work Package 2
Activity Description	Newsletters, Webinar or Digital events registrations, website management
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	It is necessary to carry out processing of common personal data for the purposes of sending newsletters, organizing webinars and other events, and for managing the website
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	We will process name, surname, age, photos, job affiliation, job role, educational background, field of expertise. We will not process any special category of data. For the website, we will process Cookie ID and IP address.
What type of data subjects will be involved?	Researchers, students, academics, general public, policy makers.
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	Potentially minors and elderly
From where will personal data be collected?	Through regular webform on the website and via cookies
What is the envisioned data flow?	The data, once submitted through the webform will be safely stored in the internal database
How will data subjects be informed about the data processing activities? (i.e., information notice)	Privacy and Cookie Policy on the website
Do you have an information notice to provide to data subjects?	It's the Privacy Policy already embedded in the website' home page.
What are the purposes of processing (i.e. why do you need the data you are collecting)?	We need the info listed in line 7 to elaborate collaterals with a dissemination's purpose.
What is the legal basis for processing?	The legal basis for processing refers to Art.6 comma a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes.
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent form presented to the data subjects	Yes, we collect consent via Privacy Form as described below. 
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	KT4D Data Controller is TRUST-IT SRL, with registered offices in Via Francesco Redi, 10 - 56124 Pisa - Italy, VAT no. and Fiscal Code IT01870130505
How long will data be stored?	Data will be stored during the project lifecycle and for five upon project conclusion, as requested from the EC. After the conclusion of the project data will be anonymised
Where will data be stored?	Data will be stored in a cloud platform, provided by AWS, located in Europe.
What organizational and technical security measures are applied to protect the personal data	<p>Organizational Security Measures:</p> <p>Access Control Policies: Strict access control policies are enforced to ensure that only authorized personnel can access the personal data located in our Cloud provider. This includes the use of strong authentication mechanisms, such as multi-factor authentication (MFA), and role-based access control to limit privileges.</p> <p>Data Privacy Policies: Organizations establish comprehensive data privacy policies that outline the handling, storage, and processing of personal data. These policies are designed to comply with relevant data protection regulations, such as the General Data Protection Regulation (GDPR)</p> <p>Technical Security Measures:</p> <p>Encryption: Personal data is encrypted both in transit and at rest. Transport Layer Security (TLS) or Secure Sockets Layer (SSL) protocols are used to encrypt data during transmission, while data at rest is typically encrypted using technologies like Amazon S3 server-side encryption or AWS Key Management Service (KMS).</p> <p>Network Security: AWS implements a robust network security infrastructure to protect personal data. This includes network segmentation, firewalls, intrusion detection and prevention systems (IDS/IPS), and distributed denial-of-service (DDoS) protection mechanisms.</p> <p>Data Backup and Disaster Recovery: Regular data backups are performed to ensure data availability and integrity. AWS offers various backup and disaster recovery services, such as Amazon S3, and AWS Backup, to facilitate reliable data protection and recovery processes.</p> <p>Logging and Monitoring: AWS provides extensive logging and monitoring capabilities to track and detect any suspicious activities or unauthorized access attempts. Services like AWS CloudWatch and AWS GuardDuty enable real-time monitoring, log analysis, and alerting.</p> <p>Physical Security: AWS datacenters are equipped with stringent physical security measures, including access controls, surveillance systems, and 24/7 on-site security personnel. These measures ensure that only authorized individuals can physically access the datacenters.</p>
Will personal data be shared with other partners and/or other third parties?	Data will be shared only with Partners and third parties listed as Data Processor in the Privacy Policy Statements.
Will personal data be transferred outside EU? If yes, please specify where.	Data are stored and managed in EU
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	No
Other relevant information	N/A

Table 22 – Trust-IT data processing activities

KT4D Partner name	University of Warwick UW
Activity Name	Tasks 4.1, 4.2, 4.3
Activity Description	Tasks 4.1, 4.2, 4.3: Developing a literature review and philosophical analysis of the values at stake in discussions about how knowledge technologies both threaten and provide opportunities to enhance democracy, including particular consideration to the use of AI to collect and process personal data, and freedom of speech.
What non-personal research data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	We will put together a number of corpora, each one dedicated to one of the topics connected with each tasks (see above). In order to do so, we will survey the existing academic literature and organise the sources into specific categories so as to make them easily searchable by the members of KT4D team.
What research data will you collect or produce?	We will collect published articles and books on the themes above.
Where will the data be stored? How will it be accessed/secured? What measures do you have in place to ensure against data loss	The lists of references created for the literature reviews will be stored in the Zotero library set up for the project. Where possible (meaning when the source is in open access or freely available), a copy of the source will also be stored in the Zotero library.
What data formats will you use? How will the data be structured/organised (metadata, filename convention,	The data collected will be stored in .pdf and .docx formats (corpora). The expected size of the data is still unknown and it will depend on how many relevant sources we will find for each different task.
From where will this data be collected? Under what kind of reuse license?	In order to survey the field and collect the academic and non academic sources needed, we will take advantage of online research repositories, library repository, but also digital libraries, websites, online newspapers, etc. in addition to printed material and resources made available by physical libraries and archives.
What if any additional measures will you take to enhance the FAIRness of the data?	We will use Zotero in order to store the list of sources (see above) as an open platform that we can make public if we choose to at the end of the project. This will make the corpora we create easy to be access and re-use by other researchers. In Zotero, we will use tags so as to make the corpora more easily searchable.
Will this/can this data be made available open access ?	It is most likely that the Zotero library containing all the references organised by area/topic and provided with the relevant tags will be freely available at the end of the project. At the moment, we are still discussing what exactly will be made openly available (if everything, or just a selection of the corpora created during the project).
Other relevant information	N/A

Table 23 – University of Warwick research data

KT4D Partner name	ICT Legal Consulting
Activity Name	Legal data protection compliance, data management plan
Activity Description	We need to provide for lawful, secure, and ethically sound data processing, sharing, and reuse in line with both current and proposed EU legislation in the project's activities.
What personal data processing activities are necessary in the context of the task(s) you are involved in? Please describe them.	We will reach out to partners of the consortium contacting them via email or phone call to facilitate communication of the tasks of the different work packages.
What personal data will you process? Will you process special categories of personal data (e.g. health data, political opinions, sexual orientation, biometric data)?	Name, surname, e-mail address, phone number
What type of data subjects will be involved?	Members of the consortium
Will you process the personal data of vulnerable people (minors, immigrants or refugees, the elderly, etc.)?	No, we will not
From where will personal data be collected?	The data collection activity will take place directly from the data subjects which are provided to the project coordinator and then shared via mailing lists and a file on Google Drive
What is the envisioned data flow?	We will refer to the personal details provided to us by the partners
How will data subjects be informed about the data processing activities? (i.e., information notice)	The project partners are aware that their name and email address need to be processed in order to carry out the activities related to the project
Do you have an information notice to provide to data subjects?	No
What are the purposes of processing (i.e. why do you need the data you are collecting)?	To facilitate communication and smooth delivery of the tasks of the different work packages in compliance with applicable data protection legislation
What is the legal basis for processing?	Execution of the consortium agreement
If the legal basis relied upon is consent, did you collect consent? If so, also provide the consent form presented to the data subjects	N/A
Who/what entity decides how personal data are processed? Please specify the name and contact details of the controller.	The project coordinator TCD determines how personal data are processed for the purpose of project administration
How long will data be stored?	Data will be stored for the shortest time possible after the project's end
Where will data be stored?	Data will be stored on Tresorit, in an ICTLC internally shared drive and Microsoft365 (emails)
What organizational and technical security measures are applied to protect the personal data	The data repository is protected by MFA and the access is restricted to those who strictly require access to the data. Email is also protected by password and MFA
Will personal data be shared with other partners and/or other third parties?	We will not share any personal data
Will personal data be transferred outside EU? If yes, please specify where.	ICTLC makes use of MS365 and Tresorit, both of which store data on servers in the EU
Will your data processing activity involve AI or machine learning techniques? If so, which model will you use and how will it be trained?	No
Other relevant information	

Table 24 – ICT Legal Consulting data processing activities

Appendix II. Personal Data Management Guidelines for KT4D Partners

Although the project is not yet actively gathering data as of the submission of this V1 of our Data Management Plan, the project partners have agreed to the following principles.

1. Always check data protection doubts you have with your organisation’s DPO!

Your DPO is best suited to help your organisation comply with the requirements under the GDPR to ensure that your organisation is not subjected to costly legal sanctions. If you don’t have a DPO, please refer to your legal department.

2. Determine the type of data being processed and continuously re-assess data being processed and the purposes of processing it

It is fundamental that your organisation maintains an Article 30 GDPR Record of Processing Activities where you map the data processing activities related to the project. ICTLC has requested that you complete a “Data Sheet” which is modelled after the Article 30 GDPR Record of Processing activities but also contains information on research data (which may or may not be personal data).

Note that there are specific requirements that must be followed for the processing of special categories of personal data.

3. Ensure you have a valid legal basis to process personal data

When consent is relied upon as the legal basis to process personal data, the KT4D Partner must ensure that such consent meets the requirements for valid consent under the GDPR. Consent must be freely given, specific, informed and unambiguous.

If you rely on consent to process personal data, you will need to keep a record of such consent. Likewise, if you rely on legitimate interest as a legal basis for processing, you will need to carry out a Legitimate Interest Assessment. Your DPO can help ensure you comply with these GDPR requirements.

4. Information and Transparency

In case of data collection, e.g., interviews, always make sure that an information notice has been provided to the data subject. Data subjects must be made aware of risks to their rights and freedoms that may be presented as a result of your processing activities. In order for them to be adequately informed, you must explicitly state everything that you will do to their data (see Articles 12-14 GDPR).

Information notices must be written in clear and plain language that takes the target audience into consideration. This means you need to ensure that it is provided to them in a language that they understand (their national language) and that translations from English versions of the information notice are accurate.

5. Rights of data subjects

Not only do you need to inform data subjects of their rights under the GDPR, you also need to allow data subjects to stop participating in the research if they no longer want to.

Note that data subject rights are not absolute rights.

If you receive a data subject rights request, be sure to immediately forward it to your DPO to handle.

6. Data Minimization

Only collect the personal data that you absolutely need to collect. If you inadvertently collect more personal data than you need, proceed to promptly delete it.

7. Data protection by design and by default

When you are determining the data processing activities that are necessary for the tasks you are working on, you must ensure that you implement appropriate technical and organisational measures to meet the requirements of the GDPR.

This means that by default, only personal data which is necessary for a specific purpose is processed.

8. Data sharing and data transfers

Ensure that you have in place the necessary contracts required under Article 28 GDPR with your data processors for any third-party services you use (i.e., Google, Microsoft, etc.).

In order to lawfully transfer personal data outside of the EEA, you must ensure that a valid transfer mechanism is used.

9. Note that anonymization and pseudonymization are two different things

It is necessary that the KT4D project partners take due care to ensure that effective pseudonymisation and/or anonymisation techniques are used when processing the personal data of research participants so as to avoid negatively impacting their rights and freedoms as a result of data processing activities in the KT4D project.

10. Data security

In addition to using a robust password, your organisation should implement multi-factor authentication on user accounts.

The strict access control policies should be enforced to ensure that only authorised personnel can access the personal data located on your cloud servers. Role-based access controls should be implemented to limit privileges.

Paper copies of documents should be stored under lock and key and kept safe from the elements (e.g., water).

11. Retention periods

You need to ensure that you establish and comply with data retention periods and have in place a procedure to delete data when it is no longer necessary.