Intelligence Network & Secure Platform for Evidence Correlation and Transfer

D8.1.0 INSPECTr Research Ethics and Data Protection

Data Management Plan

Grant Agreement No	833276	**Acronym**	INSPECTr
Full Title	Intelligence Network & Secure Platform for Evidence Correlation and Transfer		
Start Date	01/09/2019	**Duration**	36 months
Project URL	http://inspectr-project.eu/		
Deliverable	D8.1.0		
Work Package	WP8		
Contractual due date	29/02/2020	**Actual submission date**	31.07.20 (revision, living document)
Nature	R	**Dissemination Level**	PU
Lead Beneficiary	TRI		
Responsible Author	Leanne Cochrane		
Contributions from	All		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 833276.

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1		0	Initial Deliverable Structure	Leanne Cochrane
V0.2	10.02.20		First complete draft with information from partners	Leanne Cochrane
V0.3	12.02.20		Internal Review	David Barnard-Wills
V0.4	17.02.20		Peer Review inputs	Ray Genoe
V0.5	27.02.20		Peer Review and EARG feedback incorporated, LIA template added	Leanne Cochrane
V0.6	04.05.20		Updated to add CCILab/GitLab as a data storage entity.	Leanne Cochrane
V0.7	June 2020		Partner Data Tables populated.	VLTN, PHS, MOJN, LSP, GN
V0.8	24.07.20		AGS Partner Data Table populated.	AGS, TRI
V0.9	31.07.20		Review of DMP ensuring tables reflect most up to date information at date of submission post Ethics Check. All comments to partners deleted and track changes accepted.	Leanne Cochrane

Disclaimer

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the INSPECTr consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the INSPECTr Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the INSPECTr Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

Table of Contents

1 INTRODUCTION 9

1.1 Mapping INSPECTr Outputs 10

1.2 Deliverable Overview and Report Structure 11

2 OVERVIEW OF THE DATA IN INSPECTR 13

2.1 Partner Tables 14

2.2 Overview of Personal Data in INSPECTr (By Task & By Partner) 35

2.2.1 Processing Personal Data by Task for Technology Development Purposes 35
2.2.2 Processing Personal Data by Task for Human Research Participation 36
2.2.3 Processing Personal Data by Task for Dissemination 37
2.2.4 Overview of Personal Data by Partner 38

3 APPLICABLE STANDARDS, GUIDELINES AND PRINCIPLES 40

4 THE FAIR PRINCIPLES 42

4.1 The FAIR Requirements 42

4.1.1 Making Data Findable 42
4.1.2 Making Data Accessible 43
4.1.3 Making Data Interoperable 45
4.1.4 Making Data Re-Usable 45

5 PROTECTION OF PERSONAL DATA IN INSPECTR 46

5.1 Purposes and Legal Basis for Personal Data Processing 46

5.1.1 Personal data deriving from 'online sources', e.g. online image databases, online forums, dark web (ST3.2.4, ST4.4.1, ST4.4.2) 46
5.1.2 Processing LEA investigations personal data (T1.3): 47
5.1.3 Processing human research participants personal data: 48
5.1.4 Processing dissemination related personal data: 48

5.2 Data Minimisation, Storage and Retention 48

5.3 Rights of Individuals in Relation to their Personal Data and How They Can Assert Them 49

5.4 International Data Transfer 51

5.5 Partner Contacts Points and DPOs 52

6 DATA SECURITY 55

6.1 Data Security Classification 55

6.2 Data Storage and Access 55

6.2.1 OnlyOffice 55

6.2.2 CCILab/GitLab 56
6.2.3 Living Labs 56

6.3 The INSPECTr Platform 56

7 ETHICAL ASPECTS 57

8 ALLOCATION OF RESOURCES 59

9 RESPONSIBILITIES 60

10 MANAGEMENT AND COMPLIANCE 61

11 INTELLECTUAL PROPERTY RIGHTS 62

12 CONCLUSION 64

ANNEX I: INFORMED CONSENT FORMS 65

Exhibit A: Combined information sheet & information consent form 1 65

Exhibit B: Combined information sheet & information consent form 2 65

Exhibit C: Separated Information sheet and Informed consent form 66

ANNEX II: DATA MINIMISATION EXPLAINER 70

ANNEX III: LEGITIMATE INTERESTS ASSESSMENT TEMPLATE 74

List of Tables

Table 1: Adherence to INSPECTr GA Deliverable & Tasks Descriptions 11
Table 2: CCI Overview of Data Source and Scale 14
Table 3: AGS Overview of Data Source and Scale 15
Table 5: CNR Overview of Data Source and Scale 18
Table 6: EBOS Overview of Data Source and Scale 19
Table 7: EPBG Overview of Data Source and Scale 20
Table 8: GN Overview of Data Source and Scale 21
Table 9: ILS Overview of Data Source and Scale 23

Table 10: LSP Overview of Data Source and Scale 24
Table 11: MOJ Overview of Data Source and Scale 26
Table 12: PHS Overview of Data Source and Scale 27
Table 13: PSNI Overview of Data Source and Scale 28
Table 14: IGPR Overview of Data Source and Scale 29
Table 15: RUG Overview of Data Source and Scale 30
Table 16: SIREN Overview of Data Source and Scale 31
Table 17: TRI Overview of Data Source and Scale 32
Table 18: UNIL Overview of Data Source and Scale 33
Table 19: VLTN Overview of Data Source and Scale 34
Table 20: INSPECTr Tasks Processing Personal Data for Technology Development 35
Table 21: INSPECTr Tasks Processing Personal Data for Human Research Participation 36
Table 22: INSPECTr Tasks Processing Personal Data for the Purposes of Dissemination 37
Table 23: Processing of Personal Data by Partner 38
Table 24: Partner Policies on Research Data and Personal Data 40
Table 25: Table for Tracking Newly Created Datasets 44
Table 26: INSPECTr Personal Data Contact points in Partner Organisations 52
Table 27: Ethics Deliverables Due (M12-M36) 57

Glossary of terms and abbreviations used

Abbreviation / Term	Description
AGS	An Garda Síochána
BFP	Belgian Federal Police
CCI	UCD Centre for Cybersecurity and Cybercrime Investigation
CNR	Consiglio Nazionale delle Ricerche
DMP	Data Management Plan
EBOS	EBOS Technologies Ltd.
EC	European Commission
ECTEG	European Cybercrime Training & Education Group

GDPR	Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, ('General Data Protection Regulation')
GN	Ministere de L'Interieur
IGPR	Inspectoratul General al Politiei Romane
ILS	INLECOM Group
IPR	Intellectual Property Rights
LEA	Law Enforcement Agency
LED	Law Enforcement Directive (EU) 2016/680
LSP	State Police Ministry of the Interior
MoJN	Ministerie Van Justitie En Veiligheid
PHS	Norwegian Police University College
PSNI	Police Service of Northern Ireland
RUG	Rijksuniversiteit Groningen
SIREN	Sindice Limited
TRI	Trilateral Research Ltd.
UNIL	University de Lausanne
VLTN	VLTN GCV

Intelligence Network & Secure Platform for Evidence Correlation and Transfer

DATA MANAGEMENT PLAN

1 Introduction

This deliverable D8.1.0 responds to the Sub-Task 8.1.2 in work package 8 (WP8). ST8.1.2 states the following:

> *Ensure compliance to data protection requirements. Map the processing of personal data in the project – how, when, where and why it is collected, how it is processed, how it is used, stored and secured, how access is granted, mandated by the Data Management Plan, defined in WP8, to comply with the European General Data Protection Regulation (GDPR and relevant implementations of the Policing Directive (Directive (EU) 2016/680) providing that data is maintained and kept according to the regulated principles.*
> *The task links also to the activities of the IPR & Innovation Committee and follows the evolution of the DMP through the course of the project.*

The INSPECTr Data Management Plan (DMP) is stored and shared with all partners on the project's **OnlyOffice** collaboration platform.

To develop the DMP, Trilateral Research (TRI) has relied upon the EC '*Guidelines on FAIR Data Management in Horizon 2020*' and the information provided by partners to date as part of the early Ethics Requirements. The DMP is a standing item on all INSPECTr Project monthly meetings, in which TRI participates and which commenced with Technical Monthly calls in month 6. All partners have been asked to populate and keep under review their individual Tables in section 2.1 of the DMP. Task leaders should also oversee Tables 20, 21 and 22 (section 2.2) to ensure these reflect the processing of personal data by task within INSPECTr as the project develops. Table 23 is a helpful summary of the personal data processed by INSPECTr partner. All partners should inform TRI of any changes that need to be made to the DMP as soon as they become aware.

The EC suggest that a DMP should include the following:

- the handling of research data during and after the end of the project;
- what data will be collected, processed and/or generated;
- which methodology and standards will be applied;
- whether data will be shared/made open access; and
- how data will be curated and preserved (including after the end of the project).[1]

This document includes information covering all of these areas, and more. Regarding open access, the EC provides the principles of ***"as open as possible, as closed as necessary."***[2] The INSPECTr project interprets this to mean that all data generated by the project should be made open access, unless there is a pressing reason not to. INSPECTr partners will endeavour to make as much data available to the public for future research activities as is possible. It is important to be aware however, that many of the INSPECTr deliverables have 'confidential' status. This is because the tools developed, namely the INSPECTr platform, involves and targets LEAs as end-users. Due to the nature of LEA activities, conducted in the public interest, it is important that such tools are not accessible to persons engaged in criminal activity.

The INSPECTr consortium is committed to good data management and understands that doing so can improve the quality of research through generating data which is "*well organised, documented, preserved, and accessible*

[1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf p.4
[2] Ibid.

and their validity controlled at all time" can enable more efficient and excelling research not just for the consortium itself, but also for future projects.[3]

The purpose of this deliverable is to document what data the INSPECTr project will generate and how it will be used. It also explains how the use of this data will comply, as far as possible, with the FAIR principles (Findable, openly Accessible, Interoperable, Re-usable). Further, it explains how data, particularly personal data, that is generated as part of the INSPECTr project will be protected and kept secure.

This deliverable does not cover the work of INSPECTr partners away from the INSPECTr project. The entire 6-stage 'lifecycle' of data is covered in this document. The data lifecycle involves:

- **Planning research:** designing research; planning data management; planning consent for data sharing; planning data collection and processing protocols and templates; exploring existing data sources.
- **Collecting data**: data collection; capturing data and metadata; acquiring existing third-party data.
- **Processing and analysing data**: entering, digitising, transcribing, and translating data; checking, validating, cleaning, and anonymising data; creating derivative data; describing and documenting data; managing and storing data; analysing and interpreting data; producing research outputs; citing data sources.
- **Publishing and sharing data**: establishing copyright; creating user documentation; creating discovery metadata; electing appropriate access to data; publishing and sharing data; promoting data.
- **Preserving data**: migrating data to the best formats/media; soring and backing-up data; creating preservation documents; preserving and curating data.
- **Re-using data:** conducting secondary analysis; undertaking follow-up research; conducting research reviews; scrutinising findings; using data for teaching and learning.[4]
- **Disposal of data:** at the end of the data life-cycle it is necessary to dispose of data appropriately, especially if personal data.

The INSPECTr project does engage with each stage of the data lifecycle. However, the explanation of the lifecycle here does not mean that the INSPECTr partners are bound to perform each aspect of each stage in the lifecycle if these activities are unnecessary for the partners to perform their tasks.

It is anticipated that the issue of intellectual property rights (IPR) will be developed further by the Innovation Manager, Inlecom Systems (ILS) in deliverable D6.4.0 on 'Adoption and Exploitation Actions Impact Assessment and Policy Recommendations' due month 18 and through the work of the IPR and Innovation Committee (see ST8.1.2 for reference). A basic introduction to the INSPECTr approach to IPR is included in Section 11 of the current deliverable.

1.1 Mapping INSPECTr Outputs

The purpose of this section is to map INSPECTr Grant Agreement commitments, both within the formal Deliverable and Task description, against the project's respective outputs and work performed.

[3] UK Data Service, Prepare and manage data, 2020, https://www.ukdataservice.ac.uk/manage-data

[4] See UK Data Service, Research data lifecycle, 2019, https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx

Table 1: Adherence to INSPECTr GA Deliverable & Tasks Descriptions

INSPECTr GA Component Title	INSPECTr GA Component Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
D8.1.0 INSPECTr Research Ethics and data protection	*The Data Management Plan describing how research data is handled, collected, processed and/or generated by INSPECTr, for data to be findable, accessible, interoperable and re-useable (FAIR) during and after the end of the project.*	*Sections 2 - 11*	*This document sets out inter alia, the types of data collected/used during INSPECTr, how that data will adhere to the FAIR principles, details on personal data and the GDPR, and the data security measures implemented.*
TASKS			
ST8.1.2 Data Protection	*Map the processing of personal data in the project – how, when, where and why it is collected, how it is processed, how it is used, stored and secured, how access is granted, mandated by the Data Management Plan, defined in WP8, to comply with the European General Data Protection Regulation (GDPR) and relevant implementations of the Policing Directive (Directive (EU) 2016/680) providing that data is maintained and kept according to the regulated principles.*	*Sections 2 - 11*	*This document sets out inter alia, the types of data collected/used during INSPECTr, how that data will adhere to the FAIR principles, details on personal data and the GDPR, and the data security measures implemented.* *This document is the DMP.*

1.2 Deliverable Overview and Report Structure

This deliverable will be continually updated as new information is made available, with updated versions submitted as part of the subsequent deliverables on ethical governance due at months 12, 18 and 36.

INSPECTr's data management plan (DMP) will be a living document that presents the consortium's plan on the handling of research data during and after the end of the project, what data will be collected, processed and/or generated, which methodology and standards will be applied, whether data can be shared or made open access (OA)[5] and how data will be curated and preserved (including after the end of the project) in line with the H2020 Guidelines on FAIR Data Management (2016)[6]. The DMP will address the production, collection and processing of its data and scientific publications. It will provide detailed information on the project data lifecycle, privacy, and the project's policies for data collection, storage, access, sharing, protection, retention, and destruction. All consortium members will refer to this DMP if questions about INSPECTr's data policies and practices arise. TRI will maintain and update the DMP during INSPECTr's three-year lifespan, based on the input from all partners, and will monitor and report on its implementation in the project's interim and final reviews. Each project partner handling and responsible for data collected, stored or used in INSPECTr will ensure compliance with the DMP's policies and procedures, which will also support project co-ordination. Data management in INSPECTr will be respectful of and align with the project's internal ethics guidance to ensure consortium partners meet adequate ethical standards[7] and take adequate data protection measures.[8]

ST8.1.2 provides for TRI and CNR to map the processing of personal data as mandated by the DMP. This deliverable is an initial outline of our plan populated with information known to date.

[5] Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean peer-reviewed scientific research articles (published in scholarly journals) or research data (data underlying publications, curated data and/or raw data). See European Commission, H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2 21 March 2017.

[6] European Commission DG for Research & Innovation, *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, 26 July 2016. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[7] This will be informed by, among others, research integrity standards, for instance, in the ALLEA European Code of Conduct for Research Integrity. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

[8] INSPECTr will ensure compliance with Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data, and repealing Directive 95/46/EC (General Data Protection Regulation or GDPR).

2 Overview of the Data in INSPECTr

This section provides an overview of the data which INSPECTr partners expect to collect and process.

As the project is still in an early stage, this overview is provisional. It will be updated in subsequent iterations of the DMP as needed. The topic of data management is a standing item of the agenda of the monthly project meetings and will continue to be in the future. This is to ensure that TRI is kept abreast of any changes to partners plans regarding their data use and this document can be updated whenever necessary.

INSPECTr anticipates using the following **data categories** within the project**.**

- **Living labs administrative data** (participants contact details, communications from participants, communications of steering working group, identity management data and usage data on living labs Shared Space, living labs ecosystem membership data).
- **Quantitative and Qualitative evaluation data** from deployment and testing of the INSPECTr platform and tools (e.g. benchmarks, testing results, perspectives and opinions collected from end-users).
- **Evidence related data** (Universally model evidence data, and live criminal investigations data within living labs with judicial permissions) under the control and responsibility of the LEA (Law Enforcement Agencies) partners. (e.g. data for testing ingestion engines and APIs in WP3).
- **Data collected from public sources** (relevant legislation database in WP2, document database in WP3 – legislation, government guidance, codes of practices, results of ethical horizon scanning in WP8). These might provide a highly useful research resource for further exploitation and should be feasible to share widely following the project.
- **Open Source data** collected from publicly available sources (e.g. historic and real time twitter data used for machine learning technique development in Task 4.3.2). These will require ethical overview and agreement with any potential owners to arrange appropriate access.
- **Publications.** INSPECTr aims to contribute to Open Research through open peer-reviewed and other kinds of publications. As such it will produce its own interviews, reports, publications, conference proceedings, etc.
- **Dissemination related data** (The project's contact list of stakeholders, capacity building programme members, contact details for webinar and workshop participants, dissemination contact list and incoming inquiries in response to dissemination activities).

During month 6, partners have been asked to populate Table 2 which asks generally about the data collected/used in INSPECTr and specifically the **nature and scale** of the data. This table reflects **<u>all data</u>**, not only personal data (which is covered in further detail in sections 2.1 and 5 below).

2.1 Partner Tables

Table 2: CCI Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
CCI	Consortium Contact Details	xls; onlyoffice platform;	<100Mb	Project Management	Project Management	All project participants	Unclassified
	Online content from various publicly available sources. Artificial data will be used through development; e.g., fake accounts, pastes, websites, forums, etc. Only in latter stages will tool be tested on live websites. Darkweb: URLs cannot be specified as they are not fixed. (pd) Clearweb: forums include *cracked.io* and *nulled.to*. Paste sites include *pastebin.com*.	various formats	large, unknown	ST3.2.4,	To test the web scraping tool which is being designed to identify malicious activity, cross-referencing with LEA investigative data	Initially a limited number of researchers. Ultimately LEAs only	Fictitious data (dev): Unclassified Pseudonymised data: Important Pseudonymistation Keys: Confidential
	Fake accounts on social media sites, Facebook and Twitter.	various formats, texts, pictures and videos	unknown	ST3.2.4	To test the web scraping tool.	CCI Researchers	unclassified.
	Media (images, videos) from online open datasets such as *storage.googleapis.com/openimages/web/index.html* and *cocodataset.org*	any image or video format (jpeg, png, mp4, etc.	large, unknown	ST4.4.1, ST4.4.2	Training/testing computer vision applications.	A limited number of researchers.	unclassified.

Table 3: AGS Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
AGS	Closed Investigations Case Files materials, e.g.: – Seized exhibits (digital or paper) – Transcriptions of all communications, e.g. chat logs, telephone, social media, SMS (all instant messaging tools) etc. – OSINT (all available sources)	– Texts, images, pdf, doc(x), xls(x), image files, (forensic evidence files) – Doc(x), pdf – Text, image files, doc, pdf	Case dependant, variable size	Analysis work, establishing reports. In the context of T1.3 'Deployment of INSPECTr platform' by testing on real LEA data.	Verifying the suitability of the INSPECTr platform for its objective. i.e. adding efficiency to evidence gathering for the case file	LEA investigators	– Confidential – Confidential – Important
	Mocked case file data for initial CSAM use case: – Seized exhibits (digital or paper)						

	– Transcriptions of all communications, e.g. chat logs, telephone, social media, SMS (all instant messaging tools) etc. – OSINT (all available sources)						

Table 4: BFP Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
BFP	Closed Investigations Case Files materials, e.g.: – Seized documents in paper (scanned) or digital form (personal data) – Transcription of phone conversations – OSINT (publicly available data relevant to intelligence), namely facebook, linkedin, company website – Reports by investigators in the framework of the casefile	– Texts, images, pdf, doc(x), xls(x), image files, various forensic evidence files – Doc(x), pdf – Text, image files, doc, pdf, – doc(x), pdf	Case dependant, variable size	Analysis work, establishing reports. In the context of T1.3 'Deployment of INSPECTr platform' by testing on real LEA data.	Verifying the suitability of the INSPECTr platform for its objective. i.e. adding efficiency to evidence gathering for the case file	LEA investigators	– Confidential – Confidential – Important – Confidential

Table 4: CNR Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
CNR	Mobile device forensic images from the following public web sites: https://www.cfreds.nist.gov https://digitalcorpora.org https://www.droneforensics.com	Binary format (.bin) for JTAG/Chip Off acquisition; Compressed format (.tar) for UFED 4PC and (.zip) for iOS Backup.	The whole dataset, composed of 17 Android images and 18 iOS images, is about 300 GB	T2.2 and T2.4	Developing parsers to convert forensic tools XML outputs into standard (CASE ontology)	It is not data user, the only aim is to develop robust parsers being able to convert any XML or open format outputs generated by forensic tools into CASE	Unclassified, the data is not used for any kind of access

Table 5: EBOS Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
EBOS	Used:JSON of extracted rule parameters from T3.4.1.b Collected:Doc & Pdf documents including research of EU laws T3.4.1.a Output: XML (not pd)	JSON,XML, Doc,Pdf	<300MB <1GB	Task 3.4 EU Legislation Management Tools	Used by designated LEA representatives to create and input rules for cross-border cooperation and data exchange	EBOS Developers, T3.4 Partners, LEAs	Unclassified
	Used Input form various WP3, WP4, WP5 Components (not pd)	JSON XML	<300MB	Task 5.1 Generic reusable embeddable lightweight investigative widgets	This task will display inputs from various components in a light weight version	EBOS Developers, WP3, WP4, WP5 relevant partners	Unclassified

Table 6: EPBG Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
EPBG	mocked case data, fake accounts, fake/modified datasets, modified CDR/bank records etc	mostly all common image/evidence/ document formats	large/unknown	wp1	living lab testing	EPBG	confidential
	closed case data	mostly all common image/evidence/ document formats	large/unknown	wp1	living lab testing	EPBG	confidential

Table 7: GN Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
GN	Internal anonymised complaints datasets (not pd)	Text	1GB	ST4.2.1, ST4.5.3, ST4.5.4	Find typical correlations on criminal data, detection of weak signals.	Technical GN team to generate models and methods to correlate data	As we will anonymise the data is unclassified, however the dataset will not be shared or redistributed via onlyoffice. GN use only.
	Internal closed case file investigation cases - anonymised in advance. (not pd)	Text, images, pdf	4GB	ST4.5.3, ST4.5.4, ST4.3.1	Pattern detections, entity extraction, relation extraction, summarisation.	Technical GN team to generate models and methods to correlate data	Even though anonymised, regarded as confidential, GN use only, not shared or redistributed via onlyoffice.
	Darkweb crawled data - anonymised (not pd)	html, images	6GB	ST4.2.1, ST4.3.1 ST4.3.2. ST4.5.3 ST4.5.4	Pattern detections, entity extraction, relation extraction, graph relations analysis	Technical GN team to generate models and methods to correlate data.	Unclassified.
	Public language datasets E.g. http://www.statmt.org/europarl/ http://casmacat.eu/corpus/global-voices.html http://opus.nlpl.eu/UN.php (not pd)	Text	Several GB the maximum we can find during the project	T4.2	Automatic text translation	Technical GN team to generate models and methods to correlate data. Other partners who needs the dataset.	Unclassified

	Open online darkweb datasets - historic crawling on dark markets closed by LEAs E.g. https://www.gwern.net/DNM-archives https://www.kaggle.com/philipjames11/dark-net-marketplace-drug-data-agora-20142015 (not pd)	html, images	Several GB the maximum we can find during the project	ST4.2.1, ST4.3.1 ST4.3.2. ST4.5.3 ST4.5.4	Pattern detections, entity extraction, relation extraction, graph relations analysis	Technical GN team to generate models and methods to correlate data Other partners who need the dataset	Unclassified

Table 8: ILS Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
ILS	(Collected & Used) Sample Conversation Data from Public datasets e.g. Kaggle, Newcastle University (not pd)	CSV, JSON, JSON-LD, TXT	~100MB	T4.3.1/D4.3.0	Testing Natural Language Processing tools used for phone and email texts.	The T4.3.1 developers and any INSPECTr partner if they want to test/demo the tool	Unclassified
	New synthetic dataset created by INSPECTr partners to mimic conversation with hidden criminal messages. (Making a richer version of kaggle sets) (not pd)	JSON-LD, CSV	<10MB	ST4.3.1	Testing Natural Language Processing tools used for phone and email texts.	The T4.3.1 developers and any INSPECTr partner if they want to test/demo the tool.	Unclassified
	(Used, generated by LL) Extracted information from synthetic datasets (not pd)	XML, JSON, JSON-LD, RDF	10 MB - 300 GB	All WP3 tasks and deliverables, T4.3.1	Living Lab operations	Living Lab and toolbox owners will have access to the same data	Unclassified (open source)
	Possibly contact details for surveys conducted by other partners for follow up, T6.4 (non-INSPECTr LEAs and EUROPOL) (pd)	excel	<20 MB	Impact Assessments for D6.4 deliverables	Follow up for clarity on surveys etc for deliverable.	ILS.	Confidential.

Table 9: LSP Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
LSP	**Mocked data** Acquired from different test devices	various forensic evidence file formats: E01,DD/RAW,DMG,VHD,VHDX,EXT4 etc. various file formats: doc, docx, xls, xlsx, pdf, jpg, mpg, mp3,mp4, avi, bmp, ppt, pptx etc.	case dependant	analysis of electronic evidence	data acquisition testing, data analysis testing	project developers for mocked data	important
	Real data Acquired from different real case devices	various forensic evidence file formats: E01,DD/RAW,DMG,VHD,VHDX,EXT4 etc. various file formats: doc, docx,	case dependant	analysis of electronic evidence	data acquisition, data analysis	Law Enforcement for real cases	confidential

		xls, xlsx, pdf, jpg, mpg, mp3,mp4, avi, bmp, ppt, pptx etc.					

Table 10: MOJ Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
MoJN	n/a	n/a	n/a	MoJN assumes a supporting role in order to set up an information exchange infrastructure by other participants NL does not participate in actual exchange		No data will be entering or leaving NL, nor be procesed within NL	According to MS classification of participants

Table 11: PHS Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
PHS							
	Developed mock data for Terror Use Case as part of ECTEG contribution. This dataset is the IP of ECTEG but permission of use granted to INSPECTr project.	raw forensic images including all documents formats	Around 50 GB	analysis of electronic evidence D5.3, D5.4, D5.7	testing forensic tools correlation of evidence timeline	All partners	unclassified
	Reports, documentation	doc, xls, pdf, jpg	Around 2 GB	all PHS tasks	Support to the development and the reporting	Project developers	unclassified
	workshops with LEA	doc, xls, pdf	Around 1GB	all PHS tasks	Support to the development and the reporting	LEA, summary for project developers	unclassified

Table 12: PSNI Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
PSNI	Consortium contact details	xls: only office platform ;email	<100Mb	Project comms and consultation. WP involvement	Communication with other participants	2 x PSNI participants	important
	INSPECTr capabilities	docx, pdf		Defining LEA needs and seeing technical capabilities from technical partners	Outcome performance	Law Enforcement Steering Group	confidential
	Reports, publications, conference & workshop proceedings	docx, pdf,	<50Mb	Inter participant information sharing	research information sharing	Law Enforcement Steering Group	As designated on each document/output

Table 13: IGPR Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc...)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
IGPR	Acquired data from devices memory (laptop, smartphone, GPS, smartwatch and other)	Doc, docx, xls, xlsx, pdf, jpg, mpg, mp3,mp4, avi, bmp, ppt, pptx, txt and other.	Unknown	-	Link & Timeline Analysis in Digital Forensic Investigations	LEA	Confidential
	interviews, reports, publications, conference proceedings.	Docx, pdf	<10 Mb	-	Open Research	All partners	Unclassified
	Memory dumps and memory images	E01, ufdx, raw, bin and other	Unknown	-	Link & Timeline Analysis in Digital Forensic Investigations	LEA	Confidential

Table 14: RUG Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
RUG	Legislation	doc, docx, pdf	unknown	D2.1, D2.2, D3.4	Legal analysis	All partners	Unclassified
	Policies	doc, docx, pdf	unknown	D2.1, D2.2, D3.4	Legal analysis	All partners	Unclassified
	Best practices	doc, docx, pdf	unknown	D2.1, D2.2, D3.4	Legal analysis	All partners	Unclassified
	Literature	doc, docx, pdf	unknown	D2.1, D2.2, D3.4	Legal analysis	All partners	Unclassified

Table 15: SIREN Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
SIREN	CASE Standard Framework (not pd). https://caseontology.org/index.html	json-ld, json,	small.	To build the data model on which further visualisations can be built so LEA investigators can explore their own data. ST3.1.3, ST3.1.1, ST3.3.3, ST4.1.1	So as the LEA CASE information can be represented in the SIREN components of the INSPECTr platform.	SIREN researchers.	unclassified.

Table 16: TRI Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
TRI	Secondary research data on ethics standards.	Docx and pdfs	Unknown (est. <5mb)	T8.1, T8.2, T8.3	To guide ethics requirements and assessment in project, and future researchers interested in issue.	All partners. Wider public interested in D8.2.0, D8.2.1	Unclassified
	Human participant raw research data on ELSI requirements from stakeholder consultations	Docx; Audio	Unknown (word est. <1mb)	ST8.2.1	To guide ethics assessment in project.	TRI. All Partners.	Raw data – initially classified; After it depends on participant consent Outcome deliverables unclassified, i.e. D8.2.0, D8.2.1.
	Contact lists of experts and interested persons on LEAs/Tech/Ethics	Docx and xlsx	Excel est. <500kb)	T8.2	To organize stakeholder workshops.	TRI. All Partners.	Important
	Partner communications on ethics, data protection and quality issues (e.g. emails, reports, notes)	Docx or pdf	Unknown (word est. <1mb)	WP7 and WP8	To inform TRI on project tasks to enable fulfilment of INSPECTr responsibilities	TRI/All Partners	Classified

Table 17: UNIL Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
UNIL							
	Extracted information from synthetic datasets (Use Cases and test datasets)	XML, JSON, RDF	10 MB - 300 GB	Harmonisation of data formats (T2.2, T2.4)	Automated combination of data sources	Developers of INSPECTr will use data in the CASE standard to develop automated representation, translation and transformation.	Unclassified (open source)
	Legal documents	DOC, PDF	< 20 GB	T3.4.1	Extraction of named entities and information related to sharing	Developers of legal document management tool	Unclassified

Table 18: VLTN Overview of Data Source and Scale

	Description of Data Collected/Used	Data file format (e.g. doc, docx, xls, pdf, jpg, etc…)	Data file size	What work is this data needed for (data purpose)? (e.g. Task/Deliverable)	What will this data be used for (data utility)?	Whom will this data be used by (data users)?	What security classification will the data have to control access? (confidential; important unclassified) – see section 6.1
VLTN	(Collected & Used) Sample Conversation Data from Public datasets e.g. Kaggle, Newcastle University	txt pdf doc	Unknown Possibly large > 50GB	Storage Element Service T3.3.2	Testing of the storage elements - no processing of the data will be performed.	All VLTN internal researchers	unclassified
	(Collected & Used) Sample knowledge graphs taken from public sources e.g. Stanford Network Analysis Project (SNAP)	csv txt json	Unknown Possibly large > 50GB	Knowledge Graph T4.2	Testing of the knowledge graph functionality and scaling of chosen frameworks	VLTN/All partners	unclassified

2.2 Overview of Personal Data in INSPECTr (By Task & By Partner)

The data set out in the introduction to section 2 and partner tables (section 2.1) also involve the processing of some personal data. This sub-section provides a more detailed overview of the use of personal data in the INSPECTr project by task. The following information is current at m11 (July 2020) of the project, at which point the project has not processed personal data beyond consortium contact details. This section includes a timeline for when partners intend to start processing personal data subject to the legal basis.

It is presently anticipated that INSPECTr will process personal data in the following three activities:

i. personal data for technology development purposes, e.g. online sources and LEA investigations material for testing/validating tools.
ii. personal data collected from human research participants, such as for the organisation of technology tests and feedback.
iii. personal data used in our contact lists of stakeholders to whom we disseminate the results of the project.

Each category is discussed in the sub sections which follow.

2.2.1 Processing Personal Data by Task for Technology Development Purposes

Table 20 below details the technology development tasks which the consortia presently understands may process personal data. The intention is to rely on mocked/synthetic material (non-personal data structured to resemble the form of personal data) where possible. It is hoped that this will often be possible in the early stages of the INSPECTr project until the tools reach a level of maturity. In the medium to later stages of the project, to ensure the tools developed are based on realistic data, the project plans that LEA partners with Living Labs will test the INSPECTr platform on closed LEA investigation material. This will only take place in the countries where there are living labs, where the legal basis is identified in advance by LEA DPOs, and with permission from judicial authorities, where required. In addition, personal data from targeted online web pages will likely be processed for the facial recognition tool and web scraping tools developed under ST4.4.1, ST4.4.2 and ST3.2.4 respectively. See Table 20 below for detail:

Table 19: INSPECTr Tasks Processing Personal Data for Technology Development

Task	Type of Personal Data	Purpose of Processing	Starting Timeline for Processing (subject to legal basis)
T1.3	LEA closed investigations casefile data.	**Deployment of INSPECTr Platform**[9] – Continuous experimentation and integration of new components – Extended Ecosystem - Testing the tools for suitability by end users.	BFP - May 2021 All other Living Lab LEAs - May 2022
ST3.2.4	Online data - dark web and clearnet (where necessary to test after fabricated data initially used)	To develop tool capable of identifying malicious content for cross-referencing with LEA investigations data.	CCI - April 2021

[9] This task captures the processing of LEA investigations data for the testing of the tools developed in a personal data agnostic way under the other tasks.

	Darkweb URLs cannot be specified as they are not fixed. (personal data here is the bitcoin wallet when selling and buying) (highly unlikely due to sites targeted that it will include special category personal data) (nb. LEAs are not testing the web scrapper in the Living Labs, i.e. no bitcoin wallet is used to build an investigation.) Clearweb: forums include *cracked.io* and *nulled.to*. Paste sites include *pastebin.com*. (pd)		
ST4.4.1 ST4.4.2	Facial images datasets from publicly available websites. *storage.googleapis.com/openimages/web/index.html* and *cocodataset.org*	To develop facial and object recognition tools to assist LEAs with separating large forensic evidence information.	CCI - April 2020

2.2.2 Processing Personal Data by Task for Human Research Participation

In addition to personal data processed during technology development phases, the project may include the processing of personal data from human research participants outside the consortium partners, who will be asked to test the solutions on mocked data and provide feedback on issues. It is likely that only the contact information of these participants will be processed.

With regard to processing personal data for these purposes, partners will send an information sheet and informed consent form. The type of information sheet used, will depend on the level of interaction that the researchers have with the research participant. For example, it may be the case that the researchers do not engage with the participants except to send them an email or survey. In this case, the combined information sheet and informed consent form in Annex 1 (Exhibit B) can be used. On all other occasions where there is a deeper interaction between the researchers and research participant, the separate information sheet and information consent form, as set out in early draft form in Annex 1 (Exhibit C), should be used.

The tasks which may process the personal data of human research participants include those set out in Table 21.

Table 20: INSPECTr Tasks Processing Personal Data for Human Research Participation

Task	Type of Processing	Purpose of Processing	Starting Timeline for Processing (subject to legal basis)
T1.2.4	Non-INSPECTr EU LEA personnel contact details.	Stakeholder Workshops Questionnaires To review and obtain feedback from broad LEA community on the initial requirements of the INSPECTr platform. Non-INSPECTr LEAs to be recruited through the ECTEG network. (This task has however been delayed due to COVID-19 pandemic difficulty in hosting workshops)To review	CCI - August 2020

		and obtain feedback from broad LEA community on the initial requirements of the INSPECTr platform. Non-INSPECTr LEAs to be recruited through the ECTEG network. (This task has however been delayed due to COVID-19 pandemic difficulty in hosting workshops)	
T1.3.4	Non-INSPECTr EU LEA personnel contact details. (cyber and cyber assisted crime officers).	Testing mocked cases using webinars. As part of the continual iterative improvement process of the INSPECTr platform, non-INSPECTr LEA personnel asked to implement mocked use cases in workshop webinar. This webinar would likely be a progression of the ST1.2.4 workshop, utilising the same ECTEG LEA membership, as the INSPECTr tool reaches further maturity.	CCI - February 2021
ST6.3.5	Non INSPECTr LEA participation in webinars contact details.	Information and training on the INSPECTr platform. Non-INSPECTr LEAs will be recruited for participation through the CEPOL network.	CCI - September 2021
ST6.3.6	Non INSPECTr LEA participation in webinars contact details.	Information and training on the INSPECTr platform. Non-INSPECTr LEAs will be recruited for participation through the CoE C-PROG initiative.	CCI - September 2021
T6.4.3	Non INSPECTr EU LEA personnel contact details.	Surveys and Questionnaires for Impact Assessment.	ILS - March 2022
T8.2	Non INSPECTr individuals with expertise on relevant ELSI contact details.	Stakeholder workshops	TRI - August 2020

2.2.3 Processing Personal Data by Task for Dissemination

Table 22 details the tasks that are likely to involve the processing of personal data for dissemination of INSPECTr solutions. The type of personal data processed is likely to be the same as that processed in the human research participants task (section 2.2.1), i.e. the contact information of interested persons.

For developing a stakeholder contact list, partners should use the combined information sheet and consent form set out in Annex 1 (Exhibit A).

Table 21: INSPECTr Tasks Processing Personal Data for the Purposes of Dissemination

Task	Type of Personal Data	Purpose of Processing	Starting Timeline for Processing (subject to legal basis)
ST6.3.3	Names/email, contact addresses of extended INSPECTr LEA Network	Dissemination to LEA communities through existing EUROPOL and ILEANET communities.	CCI - December 2020
ST6.4.1	Names/email, contact addresses of extended INSPECTr LEA Network	Promotion of and Distribution of free INSPECTr platform to EU LEAs.	CCI - September 2020

2.2.4 Overview of Personal Data by Partner

This final sub-section of the Overview of data reflects the above information in a different form. It shows the processing of **personal data by partner**. Table 23 calls on partners to consider all their activities within INSPECTr as first identified in their Section 2.1 table and to extract from this information basic answers around personal data. In Table 23, a grey block represents a negative answer and a blue block a positive answer. This table is helpful to show that some partners do not anticipate processing *any* personal data within INSPECTr.

Please note that Table 23 refers to all personal data except INSPECTr consortium partners contact details.*

Table 22: Processing of Personal Data by Partner

	Will you process personal data during the course of INSPECTr?	**If yes, which type of processing of personal data are you involved?**	**Will it include special category data?**
CCI	Yes	Web scraping activity – dark web, paste sites, etc. (ST3.2.4) Open corpus personal data for facial recognition tool (ST4.4.1 and ST4.4.2) Contact data for dissemination activities under WP6.	Potential for special category data from online sources.
AGS	Yes	LEA Closed Investigations Data for Testing the INSPECTr platform. (relates to various tasks)	LEA case files likely to include special category data.
BFP	Yes.	LEA Investigations Data for Testing the INSPECTr platform (relates to various tasks).	LEA case files likely to include special category data.
CNR	No.	N/A	N/A
EBOS	No.	N/A	N/A
EPBG	Yes.	LEA Investigations Data for Testing the INSPECTr platform (relates to various tasks).	LEA case files likely to include special category data.

GN	No.	N/A.	N/A.
ILS	Yes.	Contact details for surveys as part of adoption and exploitation activities in WP6.	N/A.
LSP	Yes.	LEA Investigations Data for Testing the INSPECTr platform (relates to various tasks).	LEA case files likely to include special category data.
MoJN	No.	N/A	N/A
PHS	No.	N/A	N/A
PSNI	No.	N/A	N/A.
IGPR	Yes.	LEA Investigations Data for Testing the INSPECTr platform (relates to various tasks).	LEA case files likely to include special category data.
RUG	No.	N/A	N/A
SIREN	No.	N/A.	N/A.
TRI	Yes.	Contact details of ELSI experts for stakeholder workshops (T8.2)	N/A
UNIL	No	N/A	N/A
VLTN	No.	N/A	N/A

3 Applicable Standards, Guidelines and Principles

The INSPECTr project is currently drafting a project privacy policy which will be available on the INSPECTr project website from end August 2020. See, **http://inspectr-project.eu/**

In addition, partners have the following institutional policies, some of which deal with **research data** in general, and their use of personal data specifically.

Table 23: Partner Policies on Research Data and Personal Data

	Partner Policies on Processing Personal Data
CCI	https://libguides.ucd.ie/data
AGS	
BFP	
CNR	Website Privacy Policy & Cookie Policy (it): https://www.cnr.it/en/node/8446
EBOS	Website Privacy Policy: http://www.ebos.com.cy/privacy-policy
EPBG	
GN	
ILS	
LSP	
MoJN	
PHS	
PSNI	General: https://www.psni.police.uk/advice_information/information-about-yourself/data-protection/ Adult Privacy Notice: https://www.psni.police.uk/advice_information/information-about-yourself/adultprivacynoticepage/ Children’s Privacy Notice: https://www.psni.police.uk/advice_information/information-about-yourself/hildprivacynoticepage/

IGPR	https://www.politiaromana.ro/ro/legislatie/protectia-datelor-cu-caracter-personal
RUG	
SIREN	Privacy Policy: https://siren.io/legal/privacy-policy/
TRI	Privacy Policy: https://trilateralresearch.co.uk/privacy-policy/
UNIL	Website Privacy Policy: https://www.unil.ch/central/en/home/legalinformation.html
VLTN	None. Internal discussions on developing a privacy policy.

4 The FAIR Principles

The Horizon 2020 guidelines note that by making data findable, accessible, interoperable, and re-usable, this can enable data to be 'soundly managed'[10] and ultimately foster better science through enabling others to reproduce results and reuse data for future experiments.[11] This section outlines how the INSPECTr project will fulfil these principles.

4.1 The FAIR Requirements

INSPECTr partners strongly believe in the value of making research outcomes and knowledge available to the widest audiences possible, beyond project partners and user participants. The consortium will be guided in its approach by the FAIR Guiding Principles (Findable, Accessible, Interoperable, Reusable) for scientific data management. INSPECTr will respect all obligations regarding open access as described in the Grant Agreement.

In order to make data FAIR, it must be:

- Findable, meaning that the existing data is logical and that there are easy to follow rules in place that enable that data to be found.
- Accessible, meaning that as many people as possible can access the data and use it.
- Interoperable, meaning that data can easily be exchanged and used by different partners.
- Re-usable, meaning that data is licensed in such a way that future researchers can use it for subsequent research.

4.1.1 Making Data Findable

Making data easily findable is advantageous for the consortium itself as partners will be able to conduct their work in a more efficient manner. It is also beneficial for reviewers who can easily access the information they need to look at. Further, it is useful for future researchers and the public as they are more likely to be able to understand the project and use its outputs if they can easily access the different documents.

In order to make data within the INSPECTr project findable for partners during its timespan, the following measures are taken:

- **Storage Locations** – The INSPECTr consortium uses the following storage locations for data:
    - The OnlyOffice Platform for communication and storage of meeting notes, and INSPECTr partner-wide documents.
    - Living Labs dedicated servers for LEA research data, such as investigations files will be located in a managed environment, with limited access and managed control at LEA facilities. This data contains personal data and would infringe on fundamental rights to share more broadly within the consortium and beyond.
    - CCILab/GitLab Platform for INSPECTr technical partners to develop INSPECTr tools collaboratively. GitLab is a single application for the entire lifecycle of the developing technologies.
    - RocketChat will be used for all partners to aid communication within the project.

[10] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf p.3

[11] Mark D. Wilkinson et al. - The FAIR Guiding Principles for scientific data management and stewardship, Nature Scientific Data, 2016, 3(160018), pg.1

- **Naming of files** - Files are all titled according to the task or deliverable which they relate to. Partners will title documents with clear version numbers and upload date where it is necessary to distinguish similar files.
- **Documenting contributors** – All deliverables have a common frontmatter which includes a 'Revision History' table. Partners fill out this table to indicate what updates they have made to files, and when these took place.
- **Overview** - The INSPECTr OnlyOffice platform includes a 'peer review excel' document where all partners can see what documents should have been delivered, what are upcoming, and what is to be written in the future.

4.1.2 Making Data Accessible

Open access (OA) research is a major benefit for the INSPECTr project as more people will be able to access the outputs, potentially leading to greater readership and more impact. This is also required under Article 29 of the Grant Agreement, which states:

> '*Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — '**disseminate'** its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).'*

OA articles must be deposited in a repository within 6 months of publication (12 months for social sciences and humanities).[12] The partners have decided to use ResearchGate as a repository for articles, and to publish all public deliverables on the project website.

Further, Article 29 also requires: 'Each beneficiary must ensure OA (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results. 'The metadata of such publications must include the terms 'European Union (EU)' and 'Horizon 2020', the name of the action, acronym, and grant number, the publication date and length of embargo (if applicable), and a persistent identifier.[13] Additionally, partners should endeavour to have the EU emblem and 'This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833276' included in a prominent position on dissemination documents.[14]

The test data which belongs to LEAs **will not be made public** as this may contain personal data. Further, data used by LEAs will come from ongoing LEA investigations, and so is a record of a lawfully conducted invasion of privacy; releasing such data to the public would go beyond the legal mandate provided to the LEAs.

Open Access (OA) Data:

The partners aim to ensure the following data types are OA either during or after the project as appropriate.

(i) Academic Publications:

In line with the Grant Agreement, INSPECTr partners will commit to publish all academic publications (peer-reviewed publications) as OA. "Gold" OA will be targeted for all peer-reviewed scientific publications of INSPECTr. The partners have foreseen and allocated an appropriate budget for this activity. Wherever "gold" OA is not possible, "green" OA will be ensured by submitting the respective publications to, e.g., the OpenAIRE Zenodo repository. The partners prefer the gold route over the green route, because they believe the gold route will facilitate more downloads of published articles.

[12] Art.29.2, INSPECTr Grant Agreement

[13] Art.29.2, INSPECTr Grant Agreement

[14] Art.29.4, INSPECTr Grant Agreement

(ii) INSPECTr deliverables:

It is important to be aware that many of the INSPECTr deliverables have 'confidential' status. This is because the tools developed, namely the INSPECTr platform, involves and targets LEAs as end-users. Due to the nature of LEA activities, conducted in the public interest, it is important that such tools are not accessible to persons engaged in criminal activity. Other deliverables may contain information that partners intend to exploit under the IPR.

After EC approval of deliverables classified as 'public' (and not before the relevant interim and final reviews) and/or the end of the project, INSPECTr will deposit its deliverables in one or more research data repository/repositories (such as Zenodo) and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate, free of charge for any user.[15] INSPECTr deliverables will use a Creative Commons Attribution 4.0 International License.[16] According to this, a user can share (i.e., copy and redistribute the material in any medium or format) or adapt (remix, transform, and build upon the material for any purpose, even commercially, under certain conditions.

If necessary, the partners will create digested or adapted versions of the reports to specifically target intended audiences. The partners will ensure protection of knowledge by adopting licenses that enable free circulation of documents while safeguarding authors' (and the project's) intellectual rights. For instance, a creative commons license CC-BY (requiring attribution) or CC-0 (no rights reserved) license will be used for all of INSPECTr's products, to ensure that they are shared with minimal restrictions, aside from attribution to the authors or creators. Additionally, given the high value and effort that partners will spend in producing the content, they will ensure adequate protection of project and Commission's image as well as the content integrity. All publicly available project materials will be available through the project website and will be publicised through the project communication channels including newsletters and other public material, as described above. The project's Innovation Manager (ILS) will be responsible for assuring that maximum reach of knowledge is obtained, while controlling licensing and IPR issues related to content. They will report to the project management committee at its monthly meetings, once established, and will provide partners with counselling about knowledge dissemination strategies and issues.

(iii) Newly-generated Datasets:

The raw data used for training the INSPECTr platform mostly comes from openly available databases. Yet, some new data will be generated in order to improve the training of the system. These data will be made openly available after the relevant tasks have been completed, if the data provider, when available, has approved this.

Table 24: Table for Tracking Newly Created Datasets

INSPECTr created dataset	Description of contents	Owner	Dissemination and accessibility plans (e.g. link, repository entry)
Not yet created. Planned 2020.	Fake whatsapp conversations between INSPECTr partners to mimic criminal	ILS	Open AIRE AI 4 EU BDV Marketplace

[15] In line with clause 29.3 of the Grant Agreement.

[16] https://creativecommons.org/licenses/by/4.0/

	discussions for NLP testing.		

(iv) Collation of other publicly available data:

Other data collected from public sources - such as relevant legislation in WP2, the document database in WP3 which includes legislation, government guidance, codes of practices, and results of ethical horizon scanning in WP8 – could be a highly useful research resource due to the filtering and structuring to be conducted by the INSPECTr partners. Partners will aim to exploit this data by sharing widely following the project.

4.1.3 Making Data Interoperable

The INSPECTr consortium uses a variety of file formats and data sources as demonstrated in section 2.1 above. In order that documents are accessible and usable by all partners, and that documents which can be shared are accessible by the public, partners will use common file formats as much as possible. (For example .docx, .xlsx, pdf.) Such an approach is required by the partners in their use of OnlyOffice. This is software that can use common file formats in the drive itself thereby allowing all partners to contribute to and edit documents.

Further, where data is generated by the INSPECTr consortium using software that is not widely available, the respective partner will endeavour to make that data accessible with open source software.

4.1.4 Making Data Re-Usable

Some of the outputs created by the INSPECTr project will re-use existing data which is relevant to its tasks. For example, use of pre-existing data sets, or quotations from existing academic literature. Where pre-existing data is used, it will be referenced and acknowledged. Where any permissions for re-use are required, the INSPECTr partners will obtain them.

The INSPECTr consortium will ensure that any deliverables which can be made public are made public on the project website, and in open access repositories (see section 4.1.2 above on Making Data Accessible). As above, files which are open to the public will be created in a common file format.

5 Protection of Personal Data in INSPECTr

The INSPECTr project will collect and process a limited amount of personal data as is necessary for completing its goals. This section outlines how personal data which is collected and processed by the INSPECTr partners for their work in the INSPECTr project will be protected.

5.1 Purposes and Legal Basis for Personal Data Processing

The INSPECTr project will only collect personal data insofar as it is necessary to collect it for the completion of research, validation, dissemination, and exploitation of the project results. The legal basis for collection and processing of any personal data which is required for research and validation activities in INSPECTr is likely to vary, depending on the personal data and the nature of the data controller and processor.

The INSPECTr project is primarily a **research project** and to this end, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, known as the 'General Data Protection Regulation', hereafter 'GDPR'[17] is the primary basis for the data processing for technology, university, SME and LEA partners, unless the national law stipulates otherwise. **<u>It is the responsibility of the partners processing the personal data to make clear the legal basis on which they do so.</u>** They should consult their Data Protection Officers, and legal advisors.

The Ethics Manager for the project is currently collating the legal basis of all partners processing personal data, namely CCI, and the LEAs involved in the Living Labs. This process has been in motion for the past few months and involves the organisational DPOs and on occasion, the national DPAs.

This legal basis will be specified in forthcoming versions of the Data Management Plan.

Pending the receipt of information from the organisational DPOs, the following sections contain possible legal bases for the processing of personal data by data category which partners may find helpful.

5.1.1 Personal data deriving from 'online sources', e.g. online image databases, online forums, dark web (ST3.2.4, ST4.4.1, ST4.4.2)

This is the most varied data category. It is likely that some online personal data such as from online databases provide personal data based on <u>consent</u> or <u>explicit consent</u> (Article 6(1)(a) and Article 9(2)(a)). Consent should also be sought from the website hosting the data for the reprocessing in INSPECTr.

For personal data processed from other online data sources, such as online forums or the dark web, the data processors may use the legal basis of Article 6(1)(f) of the GDPR, namely that:

> (f) processing is necessary for the purposes of the <u>legitimate interests</u> pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

[17] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119/1.

Processing on the basis of legitimate interests likely requires the data processor to perform a legitimate interest assessment. A template of how to do this is included in Appendix III.

If the data processed concerns special category personal data, it may be the case that the data subject has 'manifestly made public' such data (Article 9(2)(e)). This can sometimes be difficult to ascertain and partners must evaluate whether the data subjects' 'actually intended to make their information public' if relying on this ground.[18] Data publicly available on online sources should not be considered manifestly made public by default. For that reason, partners are advised that either the explicit consent ground, or one of the two grounds below is preferred as a legal basis for processing any special category personal data.

For special category personal data, the data may otherwise be processed **only** if there are **corresponding national laws** which make this processing possible, and the activity is **proportionate** to the aim pursued. This could occur for example, under Article 9(2)(j) of the GDPR, namely that:

> (j) processing is necessary for … scientific … research 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.

It could also be asserted under Article 9(2)(g) of the GDPR, namely that:

> (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.

5.1.2 Processing LEA investigations personal data (T1.3):

In addition to the potential for processing personal data from online sources, the project plans to test the INSPECTr platform and tools on LEA investigations closed casefile data once the tools reach a level of maturity. The appropriate legal basis in this context is heavily dependent on national laws.

Where this data concerns special category personal data, under the GDPR, either Article 9(2)(j) or (g) identified above may serve as the most appropriate basis.

> (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.

Concerning Article 9(2)(j), it is of note that the recent EDPS 'Preliminary Opinion on data protection and scientific research' has stated that under this provision the 'special regime for scientific research and demonstrate that research occupies a privileged position within the GDPR.'[19] However, it also noted that the necessary member state laws 'have yet to be adopted.. [making it]… therefore difficult at present, if not impossible, to view a 'substantial public interest' as a basis for processing sensitive data for scientific research purposes.'[20]

[18] EC, *Ethics and data protection* (14 November 2018), p 13.

[19] P18, available at https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf

[20] Ibid., p23.

LEAs may further assert that this personal data is processed under the LED. Subject to a number of safeguards, the Law Enforcement Directive (EU) 2016/680[21] (LED) permits under Article 1, the:

> processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.

Special category personal data may only be processed under Article 10 of the LED where 'strictly necessary' and where similarly authorised by Union or Member State law.

The default position should be that the processing of personal data, even LEA investigations data, takes place under the GDPR.

It is possible that when testing the INSPECTr tools, LEAs may become aware of information that requires further investigative activity for legal or ethical reasons. Such material must be removed from the project sample and pursued by the LEA in accordance with the approved disclosure process to be agreed by LEAs in the INSPECTr Incidental Findings Policy.

5.1.3 Processing human research participants personal data:

The legal basis for processing personal data of human research participants asked to test or validate the INSPECTr tools is the consent of the participants (Article 6(1)(a) GDPR). In accordance with good research practice, participants will be able to withdraw their consent from the research activities, and the processing of their personal data, at any time without any negative consequences. Further, partners will, as far as is reasonably possible, endeavour to erase or rectify any data which they discover are inaccurate.[22] Please see further section 2.2.2 above and Annex 1.

5.1.4 Processing dissemination related personal data:

The legal basis for dissemination activities in INSPECTr is the legitimate interest of the partners (Article 6(1)(f) of the GDPR) in disseminating the results of their research to parties they reasonable believe would be interested in hearing about those results. Partners should carry out a legitimate interest assessment. TRI can provide assistance by way of a template for this activity should partners need. Those people contacted based upon the legitimate interests of the INSPECTr partners are free to opt out of communications at any time through using an 'unsubscribe' option which will be in plain view at the bottom of all communications these people receive. Please see Annex 1 (Exhibit A).

5.2 Data Minimisation, Storage and Retention

Partners will follow good data governance practices and will collect **no more personal data than is necessary**. INSPECTr partners will ensure that this means processing only personal data that is adequate, relevant, and limited to what is needed for their tasks. As such, partners will not collect extraneous information from

[21] DIRECTIVE (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89

[22] Art.5(1)(d), GDPR

participants where such personal data is not required to complete the task at hand. If additional personal data beyond what is necessary is provided to partners, they will destroy it as soon as is practicable.

Partners **will store personal data only for so long as it is necessary to keep it**.[23] Where identifying information is no longer needed, it will be destroyed. Partners will review the personal data they hold for INSPECTr annually starting at m12; where partners find that they are holding personal data they do not need for their future work, it will be destroyed. It is important to note, however, that the INSPECTr Grant Agreement requires partners to retain records and documentation of their work for five years following receipts of final payments from the European Commission (EC).[24] As such, personal data may be retained for this period of time where it is necessary to do so in order to provide records and documentation to the EC should they chose to carry out an audit of any partner, for example.

To ensure data minimisation, the INSPECTr partners intend to implement the following safeguards:

- The INSPECTr Platform will utilise meta-data at the Pub/Sub level;
- Data will be transformed to meet the Standardisation of Evidence Representation and Exchange (SERE) in the INSPECTr Platform thereby ensuring consistent formatting of data and disallowing non-compliant data;
- The Platform will utilise hashes and bloom filters wherever possible when developing the platform, to ensure that as much personal data as possible is obfuscated by the tool;
- Stakeholder contact lists will be organised by sub-categories (e.g. institution; purpose etc.) to minimise the storage of duplicates and outdated contact information.

In addition, investigators viewing personal data will do so on the basis of clear policies, which seek to ensure that the human investigator understands the data available as per the 'understandability' requirements set out in deliverable D9.20, including data that can be removed, and is able to make independent and objective decisions based on the data.

For **dark web personal data it will not be possible to anonymise the data**. This is because anonymisation of such data would be detrimental to the research objective, namely the cross-correlation of data to uncover criminal activity. However, dark web data is not being stored as part of the INSPECTr project, and LEA partners are not being asked to test the dark web scraper in the Living Labs.

A Data Minimisation Explainer, as shared in Deliverable D9.9, is located at Annex II to the DMP.

5.3 Rights of Individuals in Relation to their Personal Data and How They Can Assert Them

The GDPR outlines a number of rights which data subjects have with regard to their personal data. This section outlines each right and how the INSPECTr partners will endeavour to fulfil these rights as far as is reasonable.

The right to be informed: Articles 13 of the GDPR require data controllers to inform data subjects about their data processing when they collect personal data. Article 14 requires data controllers to inform data subjects about data processing when another entity acquired their data. Generally, data processors are required to

[23] Art.5(1)(e) and Recital (39), GDPR

[24] Arts.18.1 and 23.1, INSPECTr Grant Agreement

provide information on: the purpose of data processing; the retention period for personal data; who that data will be shared with.[25]

The INSPECTr partners will provide all necessary information to data subjects at the point of data collection. Where partners use data which they did not collect, they will endeavour to provide the required information to relevant data subjects unless it is impossible or would require disproportionate effort in the context of research. Where this is the case, partners will only use these data with appropriate safeguards.[26]

The INSPECTr project has a number of deliverables marked 'confidential' due to the focus of the project on developing tools to assist LEAs in conducting their lawful duties, most notably the prevention and investigation of crime in the public interest. It may not be appropriate to inform criminal suspects that their data is being processed.

The right of access: Article 15, GDPR allows data subjects to find out if their personal data is being processed, to have access to such data, and to have access to relevant supplementary information.[27]

Partners will provide this information, subject to Article 12, GDPR.

The right to rectification: Under Article 16 of the GDPR, data subjects have the right to rectify inaccurate data which is held about them, or complete data that is incomplete. This links with the requirements for obtaining accurate information noted above,[28] but requires data processers to reconsider data accuracy upon request.

INSPECTr partners will endeavour to rectify any inaccurate information which the collect and process, and will also endeavour to provide data subjects with opportunity to complete any data which is incomplete.

The right of erasure: Article 17, GDPR provides data subjects the right for their personal data to be erased from processing. This links with personal data being removed from data sets where individuals withdraw their consent, as mentioned above.

Partners will endeavour to comply with requests of erasure, but note that data processors are exempt from doing so where erasure would endanger the fulfilment of research activities (which includes safeguards to protect personal data).[29]

LEAs have processes in place to ensure the removal of entries in compliance with the law, e.g. party found innocent, legal conservation period ceased etc.

The right to restrict processing: In certain circumstances, Article 18 of the GDPR provides data subjects with grounds for restricting processing of their personal data.

Should any INSPECTr partner receive a request from an individual who wishes to restrict the processing of their personal data, they will abide by that request where Article 18(1)(a)-(d) apply and will store relevant data until the matter is resolved. Following resolution, partners will either destroy that data or process it in a way that the data subject has consented to.

The right to data portability: Article 20 of the GDPR provides data subjects with a right to request personal data which is held about them in a '*structured, commonly used and machine-readable format*' which can be used to

[25] See Arts.13 and 14, and Recitals 60-62, GDPR.
[26] Art.14(5)(b), GDPR.
[27] Meaning that information outlined in Art.1(a)-(h), GDPR.
[28] Art.5(1)(d), GDPR
[29] Art.17(3)(d), GDPR.

transfer data from one data controller to another. Data subjects can only request such data where the data is processed on the basis of consent, or a contract, and the processing is done by automated means.[30]

The right to object: Under Article 21 of the GDPR, data subjects have the right to object to processing of their personal data in some circumstances. Where personal data is collected as part of a research activity, data subjects can refuse to provide consent to processing and so this averts any need to exercise a right to object to processing.

Where personal data is processed as part of a dissemination activity, data subjects may object to processing by clicking 'unsubscribe' which will be at the bottom of all electronic communications from the INSPECTr project; this provides data subjects with an opportunity to object to processing of their personal data on the basis of the legitimate interests of the INSPECTr partners.

Rights in relation to automated decision-making and profiling: Article 22, GDPR provides people with the right not to be subject to automated decision-making or profiling which creates legal or similar effects for such persons. However, where the data subject consents to this process, such processing can be lawful.[31]

Some INSPECTr tasks involve profiling as detailed in Deliverable D9.12. It may not be possible in these tasks to inform the data subject about the existence of profiling. A number of safeguards must however be in place, such as upholding the principle of data minimisation and anonymisation where possible. Human intervention, Data Protection Impact Assessments and proportionality assessments will also be necessitated for such activities.

Data Subjects Rights will be set out in the INSPECTr Research Privacy Policy, which is due on the INSPECTr project website by the end August 2020.

5.4 International Data Transfer

It is likely that the INSPECTr consortium will transfer personal data beyond the EU. This may occur in two ways: (i) transfer of personal data to or from non-EU partners; and (ii) transfer of personal data from online sources.

It should be noted that four of the partners in the INSPECTr consortium are from non-EU countries, namely, the University of Lausanne (UNIL) based in Switzerland, the Norwegian Police University College (PHS) in Norway, Trilateral Research Ltd (TRI) and the Police Service of Northern Ireland (PSNI) both based in the UK. Transfer of personal data to or from these partners would be a non-EU country transfer. Nevertheless, it is noted that Norway is a member of the EEA and has implemented the GDPR.[32] The GDPR will also remain part of UK law for EU data subjects until the end of the transition agreement on 31 December 2020;[33] a data adequacy decision for the UK is also being sought. The EC has issued an adequacy decision in respect of Switzerland.[34]

In addition, some tasks within INSPECTr involve the collection of personal data from online sources where it may not be possible to identify where the data originates.

[30] Art.20(1)(a)-(b), GDPR.

[31] Art.22(2)(c), GDPR.

[32] The EEA has adopted the GDPR. Norway implemented the Regulation through The Personal Data Act of 15 June 2018 no. 38 relating to the processing of personal data.

[33] European Council, Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and the European Atomic Energy Community, OJ C 384 I/01 12.11.2019, Article 71 and 127.

[34] European Commission, Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland, OJ L 215/1 25.08.2000.

Where known, the INSPECTr partners commit to comply with the legislation of the country in which the data originates. Irrespective of data origin, the INSPECTr partners further commit to adhere to the Article 44 GDPR principle not to undermine the level of protection for natural persons provided under the GDPR by international personal data transfers.

5.5 Partner Contacts Points and DPOs

The INSPECTr consortium is a collection of partners each of whom are a distinct legal entity. Therefore, each partner is responsible for controlling the data they collect and process. Listed below are the persons responsible for personal data management with respect to INSPECTr at each partner, their contact details as well as the contact details for organisational DPOs, where available.

Table 25: INSPECTr Personal Data Contact points in Partner Organisations

	INSPECTr Person Responsible for Personal Data Queries	Partner DPO
CCI	Cheryl Baker Cheryl.baker@ucd.ie	Name is confidential to project. See D9.5. data.protection@ucd.ie
AGS	Tony Kavanagh tony.r.kavanagh@garda.ie	Mr Barry Lavin Data Protection Unit GDPR.DataProtection@garda.ie
BFP	Bart Swennen bart.swennen@police.belgium.eu	Michel Bruneau DPO DJSOC (Judicial Direction Serious Organized Crime) Federal Police Koningstraat 202A, 1000 Brussels michel.bruneau@police.belgium.eu
CNR	Mariangela Biasiotti mariangela.biasiotti@igsg.cnr.it	Dott. Giuliano Salberini, giuliano.salberini@cnr.it
EBOS	George Guirgis georgeg@ebos.com.cy	N/A. EBOS is not required to appoint a DPO under the GDPR and will not process personal data beyond consortium contact details. Project research privacy policy will be available on INSPECTr website from end August 2020.
EPBG	Dorel Hiir dorel.hiir@politsei.ee	Kreete Paal Kreete.paal@politsei.ee
GN	Daniel Camara daniel.camara@gendarmerie.interieur.gouv.fr	Name is Confidential to Project. See D9.5. Relying on Ministry of Interior DPO. delegue-protection-donnees@interieur.gouv.fr

		40 Avenue des Terroirs de France, 75012 Paris France
ILS	Ibad Kureshi Ibad.kureshi@inlecomsystems.com	Ibad Kureshi Ibad.kureshi@inlecomsystems.com
LSP	Olga Sernsnova olga.sersnova@ekspertize.vp.gov.lv	Ilona Reitere ilona.reitere@vp.gov.lv
MoJN	Hubb Moelker h.moelker@justid.nl	Name is Confidential to Project. See D9.5. fg@minjenv.nl Ministerie van Justitie en Veiligheid T.a.v. de Functionaris voor Gegevensbescherming Postbus 20301 2500 EH Den Haag
PHS	Yves Vandermeer yves.vandermeer@phs.no	Elisabeth Hammer Sæten Slemdalsveien 5 0369 Oslo +47 23 19 97 92 ehs@phs.no
PSNI	Graham Kissock graham.kissock@psni.pnn.police.uk	Name is Confidential to Project. See D9.5. zdataprotectionofficer@psni.pnn.police.uk
IGPR	Luiza Radu luiza.radu@politiaromana.ro	Carmen Frandes cpdpc@politiaromana.ro
RUG	Jeanne Mifsud Bonnici g.p.mifsud.bonnici@step-rug.nl>	A.R. Deenan a.r.deenen@rug.nl Office of the University Legal Affairs — Office of the University Directorate Oude Boteringestraat 44 9712 GL Groningen The Netherlands
SIREN	Jeferson Zanim jeferson.zanim@siren.io	Andrew Gallagher Andrew.gallagher@siren.io Block C 77 Sir John Rogerson's Quay Dublin D02 T804 Ireland
TRI	Leanne Cochrane Leanne.cochrane@trilateralresearch.com	N/A. Trilateral is not required to appoint a DPO under the GDPR. Organisational privacy policy is here: https://trilateralresearch.co.uk/privacy-policy/ Project research privacy policy will be available on INSPECTr project website from end August 2020.
UNIL	Eoghan Casey	External DPO relied on by University:

	eoghan.casey@unil.ch	DPO Associates SARL Place du Marché 1 CH-1260 Nyon All contacts with external DPO are to go through Pablo Diaz, <pabloandres.diazvenegas@unil.ch> at the UNIL DARIS Data and Research Information Services
VLTN	Thomas Krousarlis thomas.krousarlis@vltn.be	Veerle Leeman Director veerle.leemen@vltn.be

6 Data Security

The INSPECTr Security Advisory Board (or SAB) is responsible for information security. The Project Security Officer (or PSO) is chair of the SAB and has ultimate responsibility for information security. The PSO role is held with the INSPECTr co-ordinator and partner CCI. All partners should consult the **INSPECTr: Information Security Policy** which will be made available on OnlyOffice for detailed guidance. It is currently in draft form.

At a minimum, partners will:

- Ensure that INSPECTr research data is securely stored with clearly defined access controls (e.g. encryption (at rest and in transit), password-protection, restricting access to people working on INSPECTr);
- Ensure that INSPECTr research data is regularly backed-up;
- Ensure that local machines and servers are adequately protected in term of cyber-security (e.g. installing and updating anti-virus software, anti-malware software, and using firewalls);
- Ensure that any personal data which is processed is done in such a way as to safeguard the privacy and confidentiality of the data subject, including preventing unauthorised access or use of personal data and the equipment used for personal data processing;
- Ensure that any privacy or data protection risks associated with data processing are evaluated, and appropriate mitigation measures are in place to safeguard (personal) data.

Once the INSPECTr project has finished, responsibility for data security will lie with those managing the repositories where INSPECTr data will be available into the future.

Security classifications and repository security is briefly outlined below.

6.1 Data Security Classification

All INSPECTr data should be given a classification, according to the seriousness of the consequences of said data coming into the possession of unintended persons outside the consortium. The classifications are: (i) **confidential** if the information would cause severe reputational damage due to resultant harm; (ii) **important,** if the information would cause moderate to significant reputational damage; and (iii) **unclassified,** if the information would cause no reputational damage.

6.2 Data Storage and Access

As stated under section 4.1.1. 'Making Data Findable', INSPECTr data is stored on OnlyOffice, and on the Living Labs INSPECTr servers. In addition, partners may store certain local copies of research data on their own machines, servers, or cloud-storage systems. See Table 24 for more information on the organisational policies of each partner.

6.2.1 OnlyOffice

INSPECTr partners communicate and share information over the OnlyOffice Platform. Partners can only access the platform after being individually added by the Project Coordination team which manages the platform. After doing so, each individual must set up a personal password.

Each document shared on OnlyOffice can have different access rights. It is for the author of the document to decide on the appropriate access depending on the security classification.

OnlyOffice uses end-to-end encryption.[35]

6.2.2 CCILab/GitLab

INSPECTr technical partners will use GitLab to develop INSPECTr tools collaboratively. GitLab is a single application for the entire lifecycle of the developing technologies.

6.2.3 Living Labs

INSPECTr Project Living Lab hardware including servers, shall be located in a managed environment (fire alarm, fire suppression where possible, air conditioning, UPS, etc.), with limited access and managed access control. Servers are specifically prohibited from operating from uncontrolled cubicle areas.

Physical access to Living Labs shall be granted to the PSO and any server administrators with written approval from the PSO. Remote access to INSPECTr Living Labs must only be granted on an 'as needed' basis. The SAB/PSO will maintain an up to date list of who can access INSPECTr Living Labs.

The ability to remotely access systems will be promptly disabled when such access is no longer required.

Software will be installed only from approved internal servers to limit exposure to contaminated software. No software will be downloaded from the Internet directly to INSPECTr Living Labs.

6.3 The INSPECTr Platform

All data on the INSPECTr platform is encrypted. Two way authentication between the LEA (through IP ACL's server + user X509 certificate authentication) will take place to ensure on the appropriate end-points have access to data.

Since LEA's are required to go through a manual process to join the INSPECTr network, no unauthorised users or end-points can connect.

[35] https://www.onlyoffice.com/en/e2e-encryption.aspx

7 Ethical Aspects

Article 34 of the INSPECTr Grant Agreement requires all partners to carry out their work in compliance with ethical principles and the highest standards of research integrity. This includes abiding by 'The European Code of Conduct on Research Integrity'[36] which requires researchers to follow the principles of reliability, honesty, respect, and accountability in addition to rules on good practice and research integrity.

Any research carried out by INSPECTr partners which will involve collection of data from human participants will include the provision of information sheets and informed consent forms which abide by the standards of research ethics in addition to data protection requirements. These forms will include information on the background to and the purpose of the research, who is carrying it out, whom they can contact for further information, that their participation is voluntary and that they may leave at any time, and any risks or benefits which could be generated through their participation. Also, these forms explain in plain language what data processing may take place, and how the privacy of the participants can be safeguarded. An example information sheet and informed consent form is included in Deliverable D9.10.

Under Sub Task 8.2.1, TRI will undertake a sociological examination in discussion with work package leaders of the main ethical, legal and social issues that are relevant to INSPECTr technologies in their operational environments. The outcome of this analysis will be a set of privacy and ethical requirements in deliverable D8.2.0. due in m12 to be combined with the functional requirements. TRI has already begun this analysis as part of the Ethics Requirements deliverables that have been requested by the EC; most of which are due by m6. Based upon further analysis of the information gathered as part of these deliverables, TRI will adopt one or more methodologies to conduct this analysis over the course of the remainder of the first year of the INSPECTr project. To date, methods employed include an early project 'Ethics by Task' excel questionnaire, along with direct follow-up queries to individual partners.

The work TRI will undertake in WP8 Sub Tasks 8.2.2 and 8.2.3 will continue to inform partners of their ethical responsibilities generally and in relation to data processing, when scanning the horizon for ethical issues that could arise and when undertaking sensitization strategies for the consortium partners.

Further, under Task 8.3 TRI will adopt a responsive agile approach to assist work package 3, 4 and 5 leaders to integrate ethics and privacy by design into the INSPECTr tools.

The following table outlines the ethics deliverables due for the project:

Table 26: Ethics Deliverables Due (M12-M36)

Deliverable No.	Title	Month (M) Due
D8.1.1	First Report on Ethical Governance	M12
D8.2.0	Ethical, Legal and Social Requirements for the INSPECTr Platform and Tools	M12
D8.1.2	Second Report on Ethical Governance	M18
D8.3.0	Privacy and Ethics By Design in the INSPECTr Platform	M18
D8.1.3	Third Report on Ethical Governance	M36

[36] ALLEA, The European Code of conduct on Research Integrity, 2017, https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

D8.2.1	Ethical, Legal and Social Requirements for the INSPECTr Platform and Tools – Final Report	M36

8 Allocation of Resources

Sub Task 8.1.2 of the INSPECTr project includes the data protection requirements and the DMP and falls within the wider WP8 focus of 'INSPECTr Compliance with European Societal Values, Fundamental Rights and Applicable Legislation'.

The range of tasks within WP8 including the DMP deliverable are led by TRI. TRI has 26 person months for the total work package involving a range of significant ethical and data protection issues and requirements.

CNR is also designated as a contributor to ST8.1.2. on the data protection requirements and DMP with 2 person months allocated across this task, Sub Task 8.1.1 on Research Ethics and Task T8.3 on Privacy-by-design and Ethics-by-design for the INSPECTr tools and platform.

This document will be continually updated throughout the project as the activities and intentions regarding data usage change. A standing order of business will be made at monthly meetings for any partners who have changed their approach to using data to inform TRI so that this document can be updated.

9 Responsibilities

The planning, information gathering, and writing of this deliverable has been carried out by TRI. Each partner is responsible for providing the relevant information, and for collecting, processing, and storing data according to applicable ethical standards and legal requirements, as outlined in this document.

TRI will revise and review this document accordingly as new information comes to light during the project and will consult with partners for clarification on any outstanding issues. Revisions to this document may become necessary where: new or unanticipated datasets become available; existing datasets are re-classified due to new data protection regulations, or newly highlighted concerns; partners exit or join the consortium; technological advancements affect data security or data protection; partners update their data management, personal data, or privacy policies and this could impact on the INSPECTr project.

Partners are responsible for informing Trilateral Research of any of the above issues, or any other happenings which could have an impact on their use of data during the INSPECTr project.

Contacts that may be of use regarding the Data Management Plan include:

INSPECTr Project Manager: CCI

Viv Kearns, vivienne.kearns@ucd.ie

INSPECTr Innovation Manager: ILS

Ibad Kureshi, ibad.kureshi@inlecomsystems.com

INSPECTr Legal Manager: CNR

Mariangela Biasiotti, mariangela.biasiotti@igsg.cnr.it

INSPECTr Ethics & Quality Manager: TRI

Leanne Cochrane, Leanne.cochrane@trilateralresearch.com

10 Management and Compliance

TRI will oversee compliance with the procedures outlined in this document, alongside CCI as the project Co-ordinator. TRI will be available to partners for advice on data management throughout the project. However, each partner is directly responsible for their own compliance with the strategies and procedures discussed in this and other relevant documents.

11 Intellectual Property Rights

Intellectual property (IP) is ideas, information, and knowledge that can be traded in the form of a commodity. For example, inventions; literary or artistic works; symbols, images, and names used in commerce. Intellectual property rights (IPR) provide legal protection to the owners of IP via, for example, patents; copyrights; registered designs; registered trademarks; confidential information; database rights.

Background IP, i.e. that which is needed to implement the action or exploit the results as part of the INSPECTr project has been agreed to be provided by partners under Article 24 and 25 of the Grant Agreement.

Pursuant to Article 26.1 of the Grant Agreement, the results of the project are IP and belong to the partner who generates them. Partners who co-create results may jointly own it where they have jointly generated it and it is not possible to establish the respective contributions of each party, or it is not possible to separate them for the purposes of protecting their IP.[37] The joint owners must agree on their joint ownership in writing.[38]

Ownership of results can be transferred subject to Article 30 of the Grant Agreement. Partners must provide other partners with access to their results free of charge.[39]

Task 6.4 of INSPECTr, led by Innovation Manager Inlecom Systems (ILS), is designed to ensure that INSPECTr tools are made freely available and used by all EU LEAs, and that partners can exploit the IP generated from their work. The Innovation Manager will form part of the IPR and Innovation Committee referenced in ST8.1.2.

The partners will copyright the exploitable assets listed above under a Creative Commons Attribution 4.0 International Licence.[40] The INSPECTr consortium agrees that the publishable results of our project should be freely available, in some instances with a CC-BY-SA 4.0 Creative Commons licence, which will allow even commercial use, provided that attribution is given and the same creative commons licence is granted for further uses. Components that are already under comparable licensing will remain under their existing license, such as CASE, which uses Apache 2.0. The partners may also purchase a trademark for the technologies, tools and systems developed by the project. Trademarks help distinguish our goods and/or services from those of competitors in the market.[41]

As discussed under the FAIR requirement of 'Accessible' research, the partners will ensure protection of knowledge by adopting licenses that enable free circulation of documents while safeguarding authors' (and the project's) intellectual rights, where this does not conflict with the requirements for confidentiality set out above. For instance, a creative commons license CC-BY (requiring attribution) or CC-0 (no rights reserved) license will be used for INSPECTr's public products, to ensure that they are shared with minimal restrictions, aside from attribution to the authors or creators. Additionally, given the high value and effort that partners will spend in producing the content, they will ensure adequate protection of project and Commission's image as well as the content integrity. All publicly available project materials will be available through the project website and will be publicised through the project communication channels including newsletters and other public material, as described above.

[37] Art.26.2 and 27, INSPECTr Grant Agreement

[38] Article 26.2, INSPECTr Grant Agreement

[39] Art.31, INSPECTr Grant Agreement

[40] https://creativecommons.org/licenses/by/4.0/

[41] "A trademark helps to identify the origin of particular products and services; it creates an exclusive link between the product/service and its owner; it guarantees constant quality for the consumer; it constitutes a powerful marketing instrument." See https://trademark.eu/trademark-faq/. Two trademark service providers are Trademark.eu and the EU Intellectual Property Office (https://euipo.europa.eu/ohimportal/en).

The project's Innovation Manager (ILS) will be responsible for assuring that maximum reach of knowledge is obtained, while controlling licensing and IPR issues related to content. They will report to the project management committee (PMC) at its monthly meetings and will provide partners with counselling about knowledge dissemination strategies and issues.

12 Conclusion

This deliverable presents the INSPECTr consortium's plan to manage the production, collection, processing, and storage of data generated within the project. This deliverable will be continually updated as new information is made available, with updated versions submitted as part of the subsequent deliverables on ethical governance due at m12, m18 and m36.

Each project partner is responsible for ensuring that their handling of data within the INSPECTr project complies with the standards, procedures and strategies outlined in this document.

Annex I: Informed Consent Forms

Exhibit A is a combined information sheet and informed consent form that can be used for **inviting stakeholders to have their contact details on the project's contact list**.

Exhibit B is a combined information sheet and informed consent form that can be used in **a survey or questionnaire sent to various stakeholders where there is limited interaction between the INSPECTr researchers and respondents.**

Exhibit C is a separate and more detailed information sheet and informed consent form for **processing personal data of human research participants.**

Exhibit A: Combined information sheet & information consent form 1

Want to know more about INSPECTr?

The EU-funded INSPECTr project is creating a novel shared intelligence platform and a process for gathering, analysing, prioritising and presenting key data to help in the prediction, detection and management of crime in support of multiple agencies at local, national and international level. This data will originate from the outputs of free and commercial digital forensic tools complemented by online resource gathering. The final developed platform will be freely available to all LEAs.

The INSPECTr consortium periodically provides information about the project reports and events to stakeholders on the project's contact list. If you would like to receive such information, please reply to this e-mail and provide the following details:

Name:

Title: [optional]

Organisation:

E-mail address:

If, at any time, you would like to have your contact details deleted from the project's contact list, please send an e-mail to the Project Coordinator Cheryl Baker [Cheryl.baker@ucd.ie]. We will always offer stakeholders the option of having their contact details removed from our contact list, every time we send them any information about the project, its deliverables or events.

Exhibit B: Combined information sheet & information consent form 2

The EU-funded INSPECTr project is creating a novel shared intelligence platform and a process for gathering, analysing, prioritising and presenting key data to help in the prediction, detection and management of crime in support of multiple agencies at local, national and international level. This data will originate from the outputs of free and commercial digital forensic tools complemented by online resource gathering. The final developed platform will be freely available to all LEAs.

The INSPECTr consortium would like to have your views on INSPECTr [particular development, issue, task etc]. To that end, we would be grateful if you would respond to the following questionnaire.

In our reports to the European Commission, we will anonymise your responses, unless you tell us otherwise. As required by our grant agreement with the European Commission, we will retain your responses for five years after the EC has paid the balance of its grant to partners. Your responses will be anonymised and secured in password-protected files to which a limited number of project partners have access. The consortium backs up all files to the cloud. You may request access to your data and rectification of any your responses at any time until our deliverable is submitted to the EC and posted on the project website. You may also request deletion of your data at any time until publication of the report. The project will not share any personal data collected and processed during the project with anyone outside the consortium or the European Commission.

If you are willing to respond to the questionnaire, please provide your details here:

Name:

Title: [Optional]

Organisation:

Country:

Etc.

Exhibit C: Separated Information sheet and Informed consent form

PARTICIPANT INFORMATION SHEET

Introduction

You have been invited to take part in a research study. Before making a decision on whether you want to participate or not, **please read this document carefully**. Please ask all the questions you may have, including around risks and benefits, so you can be sure to understand all the proceedings of the study.

Description of the project

By signing the attached informed consent form, I understand that I am consenting to participate in the INSPECTr project funded by the European Union (Grant Agreement number 833276) and co-ordinated by University College Dublin. I am aware that the purpose of the activities in which I am participating is to develop a shared intelligence platform and a process for gathering, analysing, prioritising and presenting key data to help in the prediction, detection and management of crime in support of multiple agencies at local, national and international level. This data will originate from the outputs of free and commercial digital forensic tools complemented by online resource gathering. The final developed platform will be freely available to all LEAs.

The partners of the consortium are: University College Dublin, National University of Ireland (IE); An Garda Siochana (IE); Consiglio Nazionale delle Ricerche (IT); EBOS Technologies Limited (CY); European Cybercrime

Training & Education Group (BE); Politsei-ja Piirivalveamet (EE); Ministere de L'Interieur (FR); INLECOM Group (BE); State Police Ministry of the Interior (LV); Ministerie Van Justitie En Veiligheid (NL); Norwegian Police University College (NO); Police Service of Northern Ireland (UK); Inspectoratul General al Politiei Romane (RO); Rijksuniversiteit Groningen (NL); Sindice Limited (IE); Trilateral Research Ltd. (UK); University de Lausanne (CH); VLTN GCV (BE).

The INSPECTr project duration is from 01/09/2019 to 31/08/2022.

Information about your involvement

- I understand that my participation might involve interviews, workshops, webinars, surveys and/or written responses to questionnaires, where I will be invited to offer my views about [the tools developed by INSPECTr for Law Enforcement Agencies/Ethical Legal and Social issues arising from the tools planned by INSPECTr (input as appropriate)]. I am participating in these activities voluntarily, and I am free to end my participation at any time. I may refuse to answer any questions I do not wish to discuss. I understand that I have the right to ask questions and receive clear answers before making any decision. I understand that I may be asked to provide professional or personal views and that the record of my involvement in the research will be kept confidential.

- I understand that my responses to any workshop/webinar/testing lab discussion, or any interview/survey/questionnaire may be recorded and that physical copies of such recordings will be safely stored under lock and key by the INSPECTr partner leading the concerned activities. I understand that all the original data provided will be deleted five years after the project funding comes to an end, according to Article 18 of the INSPECTr Grant Agreement.

- I understand that, when the information I provide is used for the writing of the deliverable, the consortium will remove my name and all identifying features of that information so that my identity and experiences remain confidential (unless attribution is required and I have consented to it). I understand that I can request a copy of the data I have provided.

- I understand that any information that might identify me will be removed. Only the research team undertaking the research project will be able to access such data. Personal information received will be stored in separate files in a secure manner (including password protection where required). Under the General Data Protection Regulation 2016/679, the consortium has an obligation to inform me of the purpose of the collection, use, storage and retention of the information I have provided. I understand that the project will only collect information that is relevant to its activities. Personal information will be stored on internal servers, and accessible to only the partners involved in INSPECTr. The project will not transfer my personal information to third parties (i.e., people outside the project). The partners will password-protect any and all records with personal data. All computers will also have password protection to prevent access by unauthorised users. Only members of the research staff will have access to the passwords.

- I understand that this research conforms to European Commission guidelines and compliance with the current legislation.

- I understand that my responses may result in incidental and secondary findings, i.e., some information that was not the focus or primary purpose of the question(s). In such cases, I understand that I may opt out of my consent for INSPECTr's use of the incidental findings. Otherwise, I understand that INSPECTr will manage the incidental findings in the same way as the principal findings, i.e., that the information will be deleted within five years after EU project funding comes to an end and that any use of such information will be anonymised. The consortium will report incidental findings to the project's Ethics Board and, if necessary or if the Ethics Review Panel so chooses, it can evaluate incidental findings.

- I understand that I have the following rights.

My Rights	**GDPR**
I understand that I can withdraw my consent to the INSPECTr project for its processing of my personal data at any time.	Article 7
I understand that I can request access to my personal data processed by the INSPECTr project, and information about the processing.	Article 15
I have the right to receive the personal data that I have provided to the INSPECTr project in a structured, commonly used, machine readable format, where the project has processing this data in an automated way.	Article 20
I understand that if my personal data held by INSPECTr is inaccurate, I can request that it be rectified and that this amendment should be processed without undue delay. Similarly, if I feel my personal data is incomplete, I have the right to its completion and can provide a supplementary statement.	Article 16
I understand that I have the right to be 'forgotten' by requesting that my personal data be erased.	Article 17
I understand that I can request that the processing of my personal data be restricted in certain circumstances, such as where I am contesting its: accuracy; lawfulness; or that the processing of my personal data is no longer necessary for the purposes I gave my consent.	Article 18
I am aware of my right to lodge a complaint with a supervisory authority.	Article 57

- I have been given the contact details of the research team and I have been informed that I am free to contact Cheryl Baker, INSPECTr Project Coordinator, with any queries relating to my data or the project itself. The Coordinator's email address is cheryl.baker@ucd.ie.

Name:

Date: ___/___/_____

Signature:

INFORMED CONSENT FORM

Terms of consent	Respondent’s signature
I ________________ *[your name]* confirm that I have read the information sheet dated ___/___/_____ explaining the project and I have had the opportunity to ask questions about the project.	
My participation is voluntary. I agree that the data collected from me can be used for the production of an INSPECTr deliverable.	
I confirm that I agree to the activity (workshop/webinar/testing lab, interview/survey/ questionnaire) in which I am participating and that any data will only be used for the project.	
I agree to have the name of my organisation stated for this research.	
I wish for my name to be anonymised for this research. However, I give permission for members of the research team to have access to my responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the deliverable or deliverables that result from the research or are published.	
I agree to have the name of my organisation stated for this research.	
I consent to the project team contacting me, if required, as a follow-up to the research/engagement activity.	

Respondent’s name (please print):

Date:

Annex II: Data Minimisation Explainer

Data Minimisation Explainer

This document is designed to explain data minimisation in order that you can assess your work in terms of how any personal data you are using is adequate, relevant, and limited to the project. This will allow us to fulfil D9.9, ensure that the project complies with EU law, and also support later dissemination and exploitation activities. Further, by engaging in data minimisation, the likelihood of being subject to litigation for data protection failures is reduced; data cannot be abused, misused, or leaked if it is not collected, processed, or stored.[1]

Data minimisation is required under Article 5(1)(c) of the GDPR. It states:

> '*personal data shall be: …(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')*.'

The GDPR does not provide a wider explanation of this paragraph, or these terms, specifically. But, in a nutshell, the data minimisation principle requires you to identify the minimum amount of personal data which is required to fulfil the tasks you will be carrying out in INSPECTr and hold no more data than that.

This paragraph is the first of three regarding data standards. It links with requirements to ensure that personal data is accurate,[2] and is stored for no longer than necessary.[3] It also links with the principle of accountability, which requires data controllers to be able to demonstrate that personal data is collected and processed lawfully.[4] By implementing this principle in your work, and documenting it, you enable the data controllers in the project to comply with their obligations.

Data minimisation is also relevant for data processing conducted under Directive (EU)2016/680 ('The law Enforcement Directive'[5]). Article 20(1) calls upon Member state to provide (in implementing legislation) for relevant (law enforcement authority) data controllers to:

> *"tak[e] into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing, as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, both at the time of the determination of the means for processing and at the time of the processing itself, to* ***implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data protection principles, such as data minimisation****, in an effective manner and to integrate the necessary safeguards into the processing, in order to meet the requirements of this Directive and protect the rights of data subjects."*

[1] Hoepman, Jaap-Henk, Privacy Design Strategies, 2019 [p.5].

[2] Art.5(1)(d), European Parliament and Council of the European Union Regulation 2016/679 (General Data Protection Regulation) (Hereafter: GDPR).

[3] Art.5(1)(e), GDPR.

[4] Art.5(2), GDPR.

[5] Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA
OJ L 119, 4.5.2016, p. 89–131

Personal Data

Personal data is defined in Art.4(1) of the 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*.'

Therefore, personal data is that which concerns an individual and can be used to identify them. Where your actions in INSPECTr process personal data, you must engage in data minimisation.

Pseudonymisation of data may help to reduce privacy risks, but it is still personal data as it may be possible to identify the individual in question. Anonymised data is not personal data and is not subject to the GDPR. Therefore, anonymised data does not need to go through a process of data minimisation. You should be able to demonstrate how your anonymisation techniques work so that the data in question is no longer personal.

This document will now explain each of the three aspects of the data minimisation principle.

Adequate

Personal data is seen to be '*adequate*' where it is sufficient to fulfil your stated purpose; i.e. the amount of data you process is enough to carry out your tasks.[6] For example, in order to train and test a system for identifying criminals, it may be necessary to process personal data of innocent people.

Relevant

Personal data is seen as '*relevant'* where it has a rational link to your purpose; i.e. the data you process is clearly connected to the task you are performing. For example, in order to train and test a facial recognition system, it would be necessary to process images of people's faces, but details of their home life are irrelevant.

Limited

Personal data is seen a '*limited'* where it is only that which is necessary to fulfil your purpose; i.e. you hold no more data than you need to complete your task. For example, in order to train and test a voice recognition system, it would be necessary to have recordings of people's voices, but this should not mean that an excessive amount of recordings should be held and processed.

Whether the data you have is adequate, relevant, and limited must be judged on a case-by-case basis (potentially down to the individual level) and will depend upon the purpose for which you are collecting and processing the data. You must, therefore, be clear as to why you need a particular type of data before you process it.

If you are collecting or processing special category data, such as that related to criminal convictions, you will need to take special care to ensure that the data is thoroughly assessed in terms of the data minimisation principle, as processing of special category data requires extra protections for the data subject.

Implementation

In terms of implementing the data minimisation principle in the project, you first need to carefully consider what data you really need. This is not intended to constrain your actions when working on a task, but to enable you to act lawfully. Thinking about your data needs can result in designs '*that* [require] *significantly less data, or may require no personal data at all*.'[7] It also aligns with efficiency, and concerns around reducing noise in models.

[6] Information Commissioner's Office, Principle (c): Data Minimisation, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/ (Hereafter: ICO).

[7] Hoepman, n.1, [p.5].

Generally, minimisation of processing personal data can be achieved by either collecting data on less people or collecting less data about relevant people.[8] For example, if you are engaging in surveys and interview, you could consider if you need to hold and process the names of participants in these activities. Or, if you are processing data about individuals from law enforcement agencies, you could consider if you need all of the data you are provided with, and whether you could still carry out your tasks after anonymising the data. In order to engage in data minimisation, you should:

1. Determine what personal data is needed to carry out your task and process only that. You may include data that is foreseeably relevant to your purpose, but not data that has a low chance of being useful in the future.[9] You should be conservative in terms of determining what data may be necessary,[10] and should be able to justify your choices about the data you plan to use.[11]
2. Determine how long you expect you will need the personal data for prior to collecting or processing it. Protocols for deleting that data should be instigated so that the data is destroyed at the agreed time. These time frames should be reviewed periodically. They may be shortened if the usefulness of the data is expected to be completed sooner than anticipated. They should only be extended where retaining the data is necessary for completing the task. Where the data is part of a larger dataset, it may be changed to be unspecified values rather than deleted (but the entire dataset should be eventually planned for deletion).[12]
3. Exclude data that is irrelevant. Do not plan to collect or process personal data that you do not need for carrying out your task. If this data is provided to you in error, you should destroy it immediately.[13] Similarly if you find that a feature in training data is not producing any useful benefits in training an algorithm, you could minimise the data processed by deleting this feature.
4. Periodically review the data you hold, and your processing methods to ensure that they are adequate, relevant, and limited. Reviews could occur following a passage of time, for example every year, and/or after substantive events, such as after data sets have been expanded. Where data are found to be unnecessary after review, they should be destroyed.
5. Remove all personal data as soon as it is no longer useful. This should include back-ups, metadata, and traces of the data. Ensure that the data is not recoverable.[14] Note that destruction of data relates to its presence on the physical storage layer, and not just removing the data from software applications (data stripping).[15]

The UK's Information Commissioner's Office offers some guidance on data minimisation and privacy preserving techniques in supervised machine learning: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-minimisation-and-privacy-preserving-techniques-in-ai-systems/

Note that the use of techniques such as data mining, deep learning, or big data may create new personal data, or new insights into personal data. If you are using such techniques, you should be careful to ensure that only that data which is necessary is included when using these methods, and that only those insights generated from these methods that are necessary to retain are used.[16]

[8] Hoepman, n.1, [p.5].
[9] ICO
[10] Hoepman, n.1, [p.5].
[11] ICO
[12] Hoepman, n.1, [p.5].
[13] Hoepman, n.1, [p.5].
[14] Hoepman, n.1, [p.5-6].
[15] Hoepman, n.1, [p.7].
[16] Hoepman, n.1, [p.6].

Annex III: Legitimate Interests Assessment Template

INSPECTr

Legitimate interest assessment (LIA)

Template

The General Data Protection Regulation (GDPR), Article 6 provides for several possible legal bases for the processing of personal data. For example, Article 6(1)(a) allows the processing of personal data with the **consent** of the individual (the data subject). The GDPR sets a high standard for consent to be considered valid and it can always be withdrawn.

Another basis is Article 6(1)(f) which allows processing necessary for the **legitimate interests** pursued by the controller or by a third party. Our project has a legitimate interest for processing personal data, which stems from the INSPECTr Grant Agreement Article 38(1). [We are obliged to inform stakeholders, including the public, about our project, disseminate and communicate its results.]

Legitimate interest is the most flexible lawful basis for processing but must be carefully assessed in each specific case. The existence of a legitimate interest also depends on the **reasonable expectations** of the persons concerned, where we use people's data in ways they would reasonably expect and where such use would have a minimal privacy impact, and where there is a compelling justification for the processing.

In addition, a balancing exercise must be conducted between the legitimate interests of the project and the interests of the data subjects concerned. For this balancing, INSPECTr consulted the website of the Information Commissioner's Office (ICO), the UK data protection authority, which offers detailed guidance on how to conduct the balancing exercise[17] and on which the following is adapted.

The ICO says that there are three elements to the legitimate interest basis. We must

1. *identify a legitimate interest* – As stated above, we have a legitimate interest, under Article 38(1), to process personal contact details, as mentioned above, in order to create impact for our EU-funded project.

[17] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/

2. *show that the processing is necessary to achieve it* – 'Necessary' means that the processing must be a targeted and proportionate way of achieving our purpose. We must collect the contact details for LEAs across the EU in order to inform them about how our project results could help them combat crime and terrorism. The processing is necessary as we could not reasonably achieve the same result in another way.
3. *balance it against the individual's interests, rights and freedoms* – see below, where we have done a balancing test, and are confident that the consortium's legitimate interests do not override the individual's interests or fundamental rights.

The ICO posits the following questions as part of its LIA exercise. Opposite each question in the left-hand column, [we give our response] in the right-hand column.

First, Identify the legitimate interest(s).

Why do you want to process the data – what are you trying to achieve?
Who benefits from the processing? In what way?
Are there any wider public benefits to the processing?
How important are those benefits?
What would the impact be if you couldn't go ahead?
Would your use of the data be unethical or unlawful in any way?

Second, apply the necessity test.

Does this processing actually help to further that interest?
Is it a reasonable way to go about it?

Is there another less intrusive way to achieve the same result?

Third, apply the balancing test.

By applying the balancing test (based on the questions below), we consider the impact of our processing and whether it overrides the interest we have identified.

What is the nature of your relationship with the individual?
Is any of the data particularly sensitive or private?
Would people expect you to use their data in this way?
Are you happy to explain it to them?
Are some people likely to object or find it intrusive?
What is the possible impact on the individual?
How big an impact might it have on them?
Are you processing children’s data?
Are any of the individuals vulnerable in any other way?
Can you adopt any safeguards to minimise the impact?
Can you offer an opt-out?

The decision on legitimate interest

Based on the foregoing, we conclude that legitimate interest, Article 6 (1)(f), [is/is not an appropriate basis] for our processing personal data, as described above. [We are confident that our legitimate interests are not overridden by any risks to the data subject.]

Next steps

[Specify how you will inform contacts that you are processing personal data under this ground]