LAGO

Lessen Data Access and Governance Obstacles

D3.1. Consensus Report on FCT Research Landscape and Barriers for Data Sharing

Lead Beneficiary:	CENTRIC
Dissemination Level:	PUBLIC
Date:	30/04/2023
GA Number	101073951

Funded by the European Union

Project Information

Grant Agreement Number	101073951
Acronym	LAGO
Name	Lessen data access and governance obstacles
Call Topic	HORIZON-CL3-2021-FCT-01-04: Improved access to fighting crime and terrorism research data
Funding Scheme	Innovation Action
Start Date	01/11/2022
Duration	24 Months
Coordinator	ENG

Document Information

Work Package	WP3 – Framework EU FCT Trusted Research Data Ecosystem (RDE)
Deliverable	D3.1 Consensus Report on FCT Research Landscape and Barriers for Data Sharing
Date	30/04/2023
Type	REPORT
Dissemination Level	PUBLIC
Lead Beneficiary	CENTRIC
Main Author(s)	BONEVA, Gabriela MALTBY, Isaac MURRAY CANTWELL, April
Contributors	CENTRIC, CERTH, CFLW, ENG, FRMOI, IANUS, ICCS, KEMEA, LINKS, PPA, SPA.
Document Reviewers	Emmanouil Daskalakis (ICCS); Sylvie Naudet (CEA) Nikolaos Peppes (ICCS); Christos Baloukas (ICCS);
Security Reviewer	-

Revision History

Version	Date	Author	Comments
V1	17/04/2023	CENTRIC	Sent for internal review to ICCS and CEA
V2	25/04/2023	CENTRIC	Implementation of internal reviewers' comments. Version sent to Security Advisory Board
V3	05/05/2023	ENG	Security Advisory Board confirmed that no classified information is included in this deliverable. Version ready for submission

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Copyright

Abbreviations

ADR	Administrative Data Research UK
AI	Artificial Intelligence
AP4AI	Accountability Principles for Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
CCTV	Closed-circuit television
CENTRIC	Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research
CERTH	The Centre for Research & Technology, Hellas
CFLW	CFLW Cyber Strategies BV
CMINE	Crisis Management Innovation Network Europe
CSE	Child sexual exploitation
CSP	Cloud Service Provider
CTC	Cut the Cord
DCAT	Data Catalogue Vocabulary
DES	Data Encryption Standard
DHS	U.S. Department of Homeland Security
DNA	Deoxyribonucleic acid
DNN	Deep neural network
DOA	Description of the action
DOI	Digital Object Identifier
DPA	Data Protection Act
EA	Early action
EIO	European Investigation Order
ENG	Engineering Ingegneria Informatica S.p.A.
ENU	Europol National Units
EOSC	The European Open Science Cloud
EPE	Europol Platform for Experts
EU	European Union
EU IRU	Europol Internet Referral Unit
EUROPOL	European Union Agency for Law Enforcement Cooperation
EW	Early warning
FAIR	Findability, Accessibility, Interoperability, Reusability
FCT	Fight against crime and terrorism
FTP	File Transfer Protocol
GA4GH	The Global Alliance for Genomics and Health

GDPR	General Data Protection Regulation
GERE	Genomics England Research Environment
GTD	The Global Terrorism Database
HIC	Health Informatics Centre
HoG	Histogram of oriented gradients
HPC	High-performing computing
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IaaS	Infrastructure as a service
IANUS	IANUS Consulting Ltd
ICCS	Institute of Communication and Computer Systems
ICS	Customs Information System
ICT	Information and Communications Technology
IEC	International Electrotechnical Commission
IoT	Internet of Things
INTERPOL	International Criminal Police Organization
ISE	Information Sharing Environment
ISO	International Organisation for Standardisation
IT	Information technology
KEMEA	Kentro Meleton Asfaleias
LEA	Law enforcement agency
LINKS	Foundazione LINKS - Leading Innovation & Knowledge for Society
ML-DL	Machine Learning – Deep Learning
MOD	Ministry of Defence
MOI	Ministry of the Interior
NATO	The North Atlantic Treaty Organisation
NHS	National Health Service
NIFO	National Interoperability Framework Observatory
NL	Netherlands
OCRE	Open Clouds for Research Environment
ONS	Office for National Statistics
OSINT	Open-source intelligence
OSP	Online Service Provider
OSPP	Open Science Policy Platform
PaaS	Platform as a service
PPA	Politsei- ja Piirivalveamet
R&D	Research and Development
RDA	Research Data Alliance

RDE	Research Data Ecosystem
RDF	Resource Description Framework
RSA	Rivest-Shamir-Adleman encryption
SaaS	Software as a service
SCSN	Scottish Community Safety Network
SFTP	Secure File Transfer Protocol
SIS	Schengen Information System
SIS II	Schengen Information System II
SPA	Polichmyndigheten Swedish Police Authority
SPN	SafePod Network
TFTP	Terrorist Finance Tracking Program
TRE	Trusted Research Environment
TRUST	Transparency, Responsibility, User focus, Sustainability, Technology Principles
UK	United Kingdom
UKHDRA	United Kingdom Health Data Research Alliance
UKRI	United Kingdom Research and Innovation
VIS	Visa Information System
VML	Virtual Microdata Laboratory
VPN	Virtual Private Network

List of Figures

List of Tables

Table of Contents

Executive Summary

This report has been prepared as part of WP3 (Task 3.1: Analysis of FCT research data landscape and gap analysis). T3.1 is a crucial first step in illustrating the current state of the FCT research landscape and providing a fully comprehensive blueprint for all stakeholders involved. To achieve this end this report in particular refers to a broad array of pertinent topics such as current practices, strategies, initiatives, tools, and resources via a detailed literature review and written expert consultations. In doing so, it provides a detailed overview of the gaps, challenges, barriers, and enablers to the creation of an FCT research data ecosystem.

The primary objective of this report is to concretely identify the barriers in FCT research data sharing, map existing data initiatives and projects and therefore build a broad consensus to address the notable lack of trust between different organisations and stakeholders. It is written with the intention to provide stakeholders within the LAGO consortium a detailed analysis of the current operational parameters of the FCT research Landscape.

It is intended that this report will facilitate in clarifying the requirements for the LAGO Reference Architecture in the creation and provision of research data, the usage and exploitation of such data and finally the governance of the ecosystem and its data spaces.

The findings from this report will serve as the foundation towards the creation of a comprehensive roadmap that will enable the implementation of a trusted EU FCT Research Data Ecosystem that also has the potential to be applied to other fields outside of fighting crime and terrorism. The roadmap will consist of step-by-step guidelines, phases, stages, and steps including "procedural know-how" for the successful implementation of the FCT Research Data Ecosystem.

1 Introduction

1.1 Background and Objectives

This report is part of LAGO's Work Package 3 "Framework EU FCT Trusted Research Data Ecosystem (RDE)", which aims to analyse existing data systems, sources, and barriers to the adoption of an EU-level research data ecosystem (RDE). Additionally, this WP aims to define use cases and identify pre-requisites and requirements for the establishment of the data ecosystem. This will support the LAGO project to design an effective reference architecture and to provide a reference implementation of a software framework enabling the envisioned ecosystem, including integration, deployment, and access aspects.

T3.1 is specifically aimed at the analysis of the fight against crime and terrorism (FCT) research data landscape and gap analysis. This includes reviewing existing research data landscapes to understand current practices, resources, tools, gaps and barriers. To gather this information, the task has:

1. Conducted an analysis of reports, guidelines and other documentation that address data sharing in applied research (FCT specific, as well as other complex areas such as health).
2. Conducted elicitation exercises with relevant stakeholder groups to identify best and effective practices, current and future usage scenarios, requirements and factors that impact ability or willingness for data sharing.
3. Analysed existing data systems and sources relevant to FCT research, data formats and mechanisms for data transfer, storage and security.
4. Identified research and development (R&D) projects for relevant datasets and methods for creating them (e.g., AIDA, GRACE, STARLIGHT).
5. Gathered perspectives from Police directive, Europol directive, Frontex and legal framework in EU MS.

1.2. Deliverable Structure

The deliverable is divided into *six sections* to present the different types of data collected. **The present section** explains the objectives of Task 3.1 and the structure of the report.

The Second Section explains the bi-phased methodology that was utilised to compile all the required information for the successful achievement of the goals of Task 3.1. The methodologies applied during the desk research phase and the expert consultation phase are described.

The analyses of the surveys focus on four main topics: *i)* existing practices, initiatives, and data strategies in the domain of FCT RDE; *ii)* identifying the tools and resources required in FCT research, and the barriers and enables in the domain; *iii)* analysing the good practices and required risk assessments and management in FCT RDE; and *iv)* mapping future requirements, including gaps and opportunities in FCT research and data sharing.

The Third Section provides the results of extensive desk research concerning existing projects and initiatives that are relevant to the scope of this research.

The Fourth Section provides a detailed overview of current data practices, tools, and regulations in the domain of research dataset ecosystems (RDE). The main objective of this section is to present the practices involved with the local and transnational sharing of sensitive data, and the utilised data tools and regulations concerned with sensitive data sharing and processing. Hence, the third section introduces best practices and lessons learned that can be applied in the development of the fight against crime and terrorism (FCT) RDE framework of LAGO. This section offers a multispectral perspective of the current data practices by covering different geographical areas.

The Fifth Section refers to data practices which are relevant to the requirements of building a successful FCT RDE.

The Sixth Section details the analyses garnered from the in-depth expert written consultations with leading EU experts in FCT research. It details key enablers, barriers and risks which will either hinder or support the adoption of the RDE. These enablers, barriers and risks related to key areas such as technology, data, organisational, economic, professional, cultural, governance and policy considerations.

The Seventh Section assesses the gaps in knowledge relating to the FCT research domain. It refers to a lack of available knowledge in the following areas: best practices, infrastructure and resource requirements, technology, policy, governance, cultural and organisational gaps.

The Eighth Section refers to recommendations and lessons learned from consulted experts which encompass effective data handling practices, recommended data standards and policies, best practices in FCT research, finance, competencies, and training provision.

The Ninth Section includes the conclusions of the information explored in the previous sections, along with elaboration on research limitations.

2 Methodology

In order to gain holistic insight into the current FCT landscape, (the challenges, gaps, and requirements) a two-tiered approach was developed. The development of T3.1 was carried out using two methodological tools:

1. A comprehensive desk-based literature review examining a variety of documents relating to FCT research activity such as *reports, guidelines, data strategies, initiatives,* and *policies*. Reviewing these assisted in the development of part two of our research activity;
2. The written expert consultations. A template was developed for dissemination to experts identified by our partners.

The comprehensive literature review involved extensive desk-based research but also involved T3.1 partners IANUS, PPA, CERTH, ICCS, KEMEA, ENG, LINKS, CFLW and SPA. A methodological framework for T3.1 (inserted into Appendix A – Methodological Framework) was developed by CENTRIC as task leader, having shared with all ten partners at the WP3.1 kick off meeting in January 2023.

2.1 Desk Research

The first research activity within T3.1 was conducted by CENTRIC. Initial searches were conducted by the CENTRIC team using the following search portals: Google Scholar, Scopus, Sheffield Hallam Library and OpenAIRE. This activity provided an overview of the types of documents available and played a fundamental role in creating a comprehensive data summary review template document for dissemination to partners. The purpose of this template was to facilitate partners across various degrees of FCT experience to identify and summarise documents we identified as relevant to our research query.

The desk research on existing data practices within the FCT research domain was extensive and included input from 11 of the partners involved in T3.1. An extensive analysis utilising the document summary review template provided by the lead partner CENTRIC, facilitated a solid foundation at which to build a full spectrum view and blueprint of the current state of FCT research data sharing. As part of this research seventy-eight documents were summarised and referred to strategy, data sharing practices and infrastructure reports.

The document summary review template (Appendix B – Document Summaries Template) was then disseminated to our ten partners requesting them to effectively summarise 10 documents of their choosing according to the literature inclusion criteria. To avoid duplication of efforts a Master excel was created for the partners to refer to prior to undertaking their document summary review. Desk research was used to provide a comprehensive overview of the current state of the fighting crime and terrorism research landscape. Furthermore, it was used to assist in the development of data collection tools (e.g., written expert consultations) to identify areas in which the literature does not cover sufficiently in contextualising and illustrating the state of research landscape as it currently stands.

The template included a literature inclusion criterion with guidance on the age of the documents (2018 onwards) to guide partners as to the type of documentation our research inquiry required (please see Appendix B – Document Summaries Template).

Table 1: Literature Inclusion Criteria

Literature Inclusion Criteria	
Please note that we are seeking documentation that refers to the following topics:	
Time	**from 2018 or later (after GDPR introduction); older documents if they are providing information which is generally relevant/not impacted by GDPR**
Domain	**FCT research; also documents from research domains with similar complexity of data production, data sharing etc. (e.g., health, military)**
Document types	**all types - examples see below**
1	Existing and emerging National Data and EU Strategies
2	Current infrastructure, infrastructure changes and tools
3	Methods on risk and impact assessments
4	Existing initiatives and standards for the creation of Data Spaces across the EU
5	EU training and testing data ecosystem for FCT research
6	Current governance structures with specific roles as well as risk-aware guidelines based on FCT horizontal issues such as crime type or domain (e.g., Cyber Crime, terrorism, CSE, firearms and other illegal trafficking).
7	Impact and policies
8	Open specifications and standards
9	Standards to access quality, interoperable and portable data (and tools)
10	EU Data Strategy
11	Regulation on National/European data governance (e.g. regulations specific to counter terrorism, etc.)
12	Challenges including societal sensitivities, cross-border dimensions, disparate purposes of data sharing:(e.g., for training and testing of AI or Big Data solutions) as well as potentially conflicting interests (e.g., between LEAs, researchers, and industry), elevated quality requirements of datasets as well as risks and issues of proportionality.
13	Handbooks, manuals, training materials
14	Whitepapers, industry reports, reports by data providers or similar
15	academic articles, books, chapters, conference papers
16	Other (please specify)

Collectively, over 80 resources (i.e., articles, books reports, official websites etc) relating to the FCT research activity as it currently stands were located and summarised.

2.2 Expert Consultations

This activity within T3.1 consisted of using the DOA work package objectives and comprehensive literature review to compile relevant and probing questions for experts. These questions were to gain a deeper appreciation of the key barriers and enablers to the creation of an FCT RDE. The written expert consultations were used to collect data from stakeholder groups across various areas of specialisation such as CT (Counter Terrorism), CSE (Child sexual exploitation), Public space protection, crisis management and infrastructure protection. Stakeholders were recruited via partner networks within Task 3.1 and came from a policymaker, LEA, industry, or academia background.

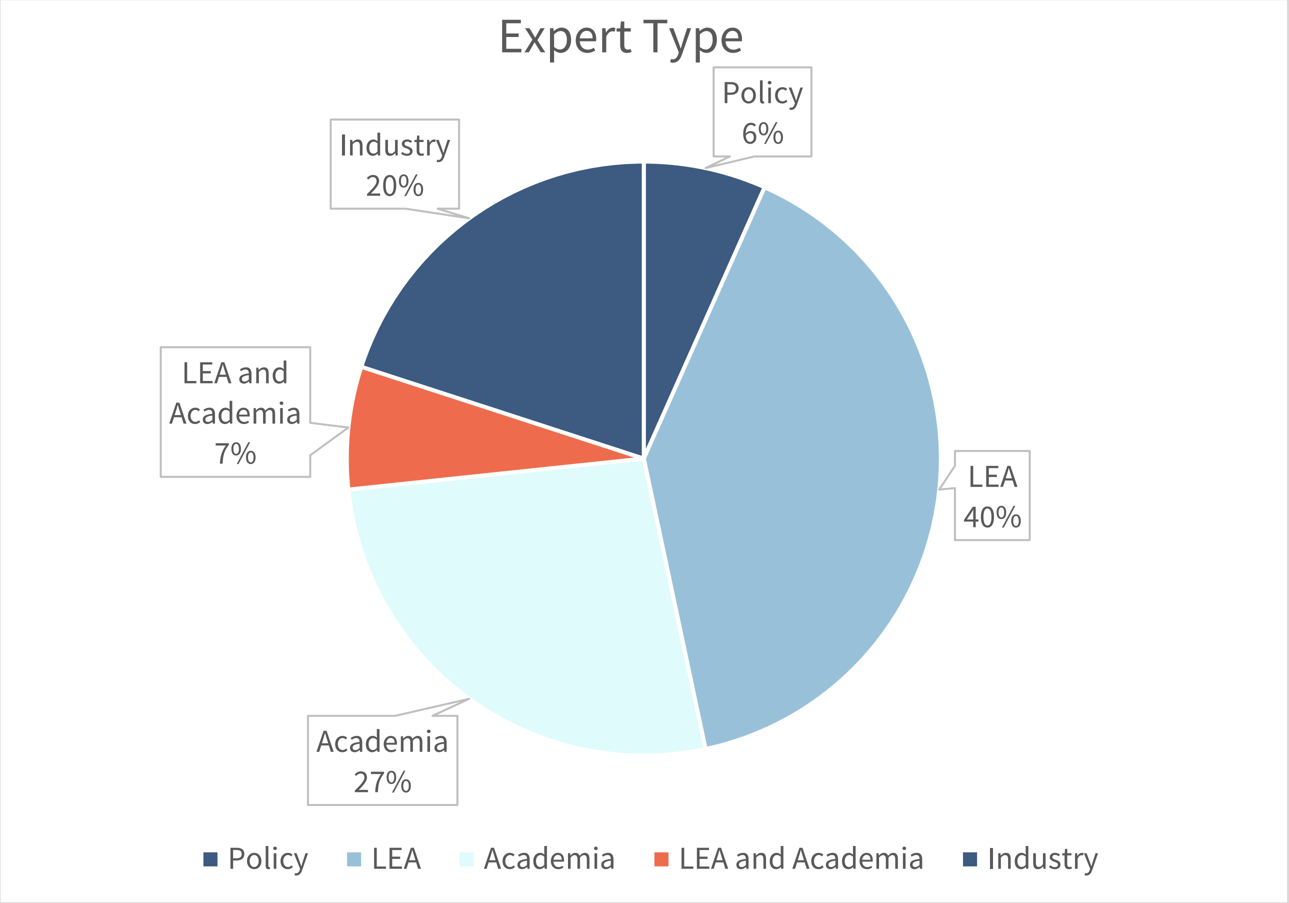


Figure 1: Type of experts consulted.

For recruitment partners sent identified experts in their network a prewritten recruitment email detailing the specifications of the project and what the research activity entailed. Once the identified expert confirmed their interest in participating, the written expert consultation form including the consent form was forwarded to them for consideration and completion.

The written expert consultation was split into four sections (for the detailed guideline see Appendix C – Written Expert Consultation Questions):

Section 1: Current State of Practice

- Current and existing initiatives and data strategies
- Current usage scenarios (feeds into T3.2)
- Current existing standards and policies
- Data types and sharing in the RDE.

Section 2: Tools, Resources, Barriers, and Enablers

- Tools to support FCT Research
- Resources for FCT Research
- Most important barriers and enablers (5)
- Trust in FCT research

Section 3: Good Practices including risk assessment and management.

- Risk assessment and management
- Good Procedures and Practices

Section 4: Mapping Future Requirements

- Opportunities and Gaps for future research and development
- Invitation to provide further aspects/insights.
- Invitation to provide additional consultations.

The type of information collected via the written expert consultation forms involved information on their level of expertise in FCT research, the type of organisation they work for and their area of research. Experts were then asked a variety of targeted questions examining the current state of practice in FCT research, tools, resources, barriers, and enablers of FCT research, good practices including risk assessment and management and finally questions on what future requirements a successful FCT RDE would need to incorporate.

All the data collected from these consultations was input into excel for thematic analysis by CENTRIC's team to determine key projects, initiatives, standards and priorities as identified by experts.

Sample description: Countries, level of FCT research experience and background

Out of the 30 completed written expert consultations 23% had limited FCT research experience, 23% had moderate FCT experience with 46% had very high FCT experience with 7% not elaborating on their FCT experience.

Limited experience despite having knowledge of research in this area might denote that more adequate training is needed to provide a deeper awareness around data sharing procedures within this particular field.

Table 2: Countries covered through the expert consultations.

Country	Completed Expert Consultations
UK	12
Greece	8
Netherlands	2
Italy	4
Estonia	4
	30

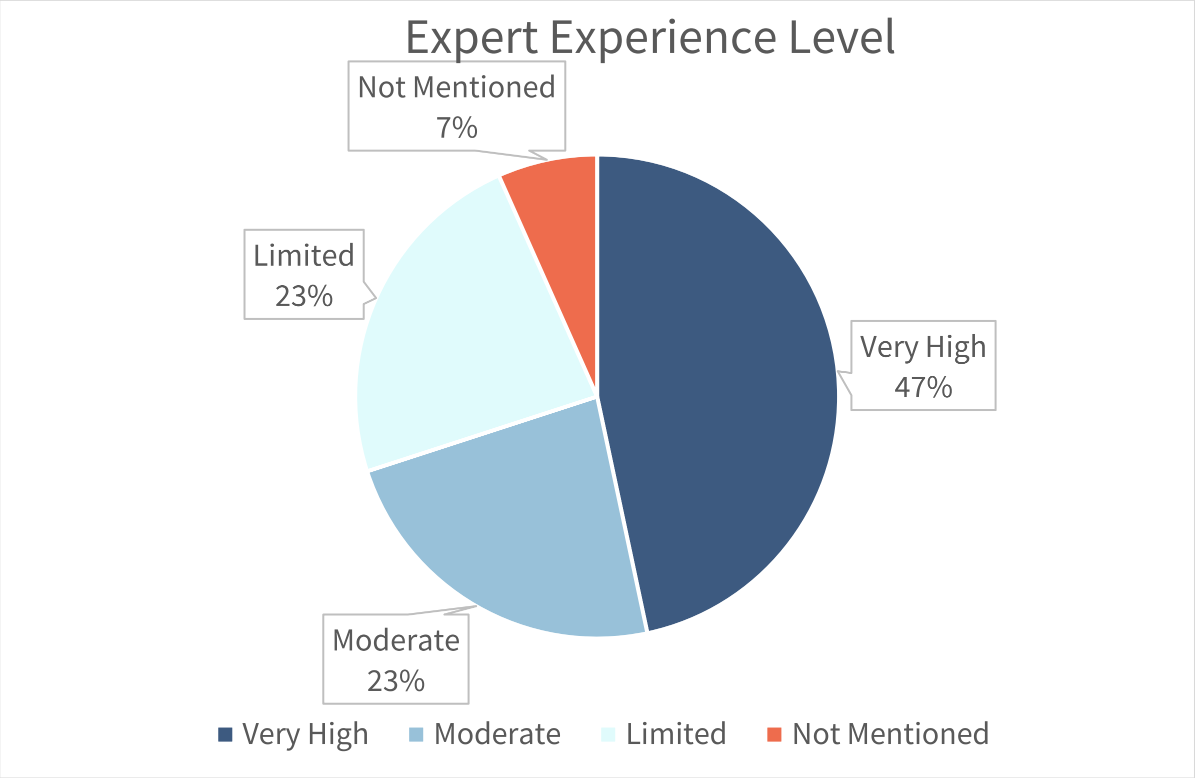


Figure 2: Experience level of the experts consulted.

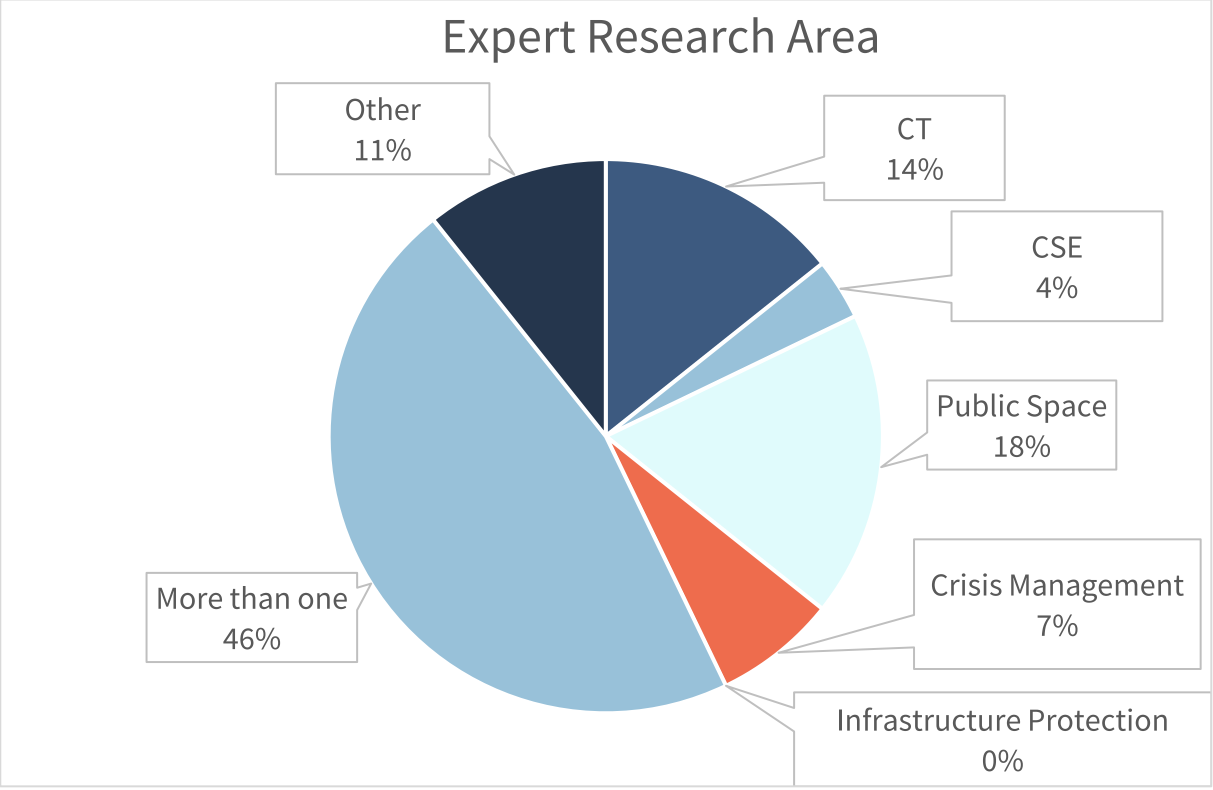


Figure 3: Domain background of the experts consulted.

3 Overview of Relevant Projects and Initiatives

The following tables (Tables 3-5) provide an overview of the projects and initiatives experts considered relevant and of interest for consultation and potential learnings in the creation of the LAGO RDE. Where possible websites to the projects are provided together with a short explanation.

Table 3: Initiatives named by the experts relevant to RDE in FCT Domain

Named Initiatives		Notes
1	**The European Interoperability Framework**	https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail
2	**EU innovation lab under EUROPOL**	Developing a sandbox/data space for RDE purposes: https://www.europol.europa.eu/operations-services-and-innovation/innovation-lab
3	**EU-LISA**	Responsible for developing EU large scale information systems in the field of justice and Home Affairs and their interoperability. https://www.eulisa.europa.eu/
4	**The Crime Open Database**	(CODE; outside the EU, data on crime from multiple larger US cities) https://osf.io/zyaqn/
5	**The Global Terrorism Database (GTD)**	At the international level, e.g., (Open pseudonymised survey data that includes questions about trust in the police, perceived security in the home neighbourhood, attitudes to criminal behaviours, etc. https://www.start.umd.edu/gtd/
6	**European Social Survey**	(ESS; comparative data collected in European countries) (WVS; data collected in countries around the world). Sources of aggregated statistics on crime, and drug use. https://www.europeansocialsurvey.org/
7	**World Value Survey**	https://www.worldvaluessurvey.org/wvs.jsp
8	**Eurostat statistics on crime and criminal justice**	(European countries). https://ec.europa.eu/eurostat/web/crime
9	**EMCDDA statistical**	Tables on the various aspects of drug use (European countries). https://www.emcdda.europa.eu/data/stats2022_en
10	**Schengen Information System (SIS)**	This information system is shared among Schengen countries and allows law enforcement authorities to share data regarding wanted persons, stolen objects, and other information relevant to criminal investigations. https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/schengen-information-system_en
11	**Schengen Information System II (SIS II)**	SIS II is a centralized database used by law enforcement authorities in EU countries to share information on people and objects of interest in the context of border checks, visa and residence applications, and law enforcement activities. It is a critical tool in the fight against terrorism, organized crime, and other serious offenses. https://knowledge4policy.ec.europa.eu/dataset/ds00009_en
12	**Customs Information System (ICS)**	This information system is shared among EU countries and allows customs authorities to share information about the trade of goods and other relevant data to investigate customs crimes. https://taxation-customs.ec.europa.eu/customs-4/customs-security/import-control-system-2-ics2-0_en
14	**Visa Information System (VIS)**	This information system is shared among EU countries and allows visa authorities to share information about visa applicants, including biometric

		data, to prevent fraud and false identity. All these initiatives are related to criminal investigation and security in the EU, and all involve sharing data among law enforcement, customs, or judicial authorities. Therefore, they can provide models for creating an RDE exclusively for the fight against crime and terrorism.
15	**European Investigation Order (EIO) System:**	This system is a network of national authorities and databases that enables the exchange of information and evidence between EU countries in criminal proceedings. The system provides a standardized procedure for obtaining evidence across borders, facilitating cross-border cooperation in the investigation and prosecution of crimes.
16	**Terrorist Finance Tracking Program (TFTP):**	This joint program between the EU and the United States facilitates the exchange of financial information to prevent terrorism and terrorist financing. The program allows EU and US authorities to share information on financial transactions that may be linked to terrorist activities. Passenger Name Record (PNR) system: This database collects information about passengers traveling by air to or from the EU. The system is used by law enforcement authorities to identify and track individuals who may pose a security risk, including terrorists and other criminals. Integrated Police Cooperation Application (IPCA): IPCA is a web-based platform that allows law enforcement authorities in EU countries to share information and intelligence related to cross-border crime, including terrorism, organized crime, and cybercrime. The platform provides real-time access to shared databases, enabling faster and more effective cooperation among law enforcement agencies. These projects illustrate the significance of data sharing and collaboration in the fight against crime and terrorism. By enabling law enforcement authorities to access and share information across borders and jurisdictions, these databases and systems help to strengthen Europe's security and prevent serious crimes.
17	**The European Open Science Cloud (EOSC):**	The EOSC is a pan-European initiative that aims to provide a virtual environment for researchers to store, manage, and share data across disciplinary and national borders. It supports open science and aims to provide a user-friendly infrastructure for research data.
18	**The Global Alliance for Genomics and Health (GA4GH):**	GA4GH is an international alliance that aims to promote the sharing of genomic and health-related data in a secure and ethical manner. It develops standards and frameworks for the responsible sharing of genomic and health-related data. The FAIR Data Principles: The FAIR data principles aim to ensure that research data are Findable, Accessible, Interoperable, and Reusable. They provide guidelines for making research data more discoverable, accessible, and reusable.
19	**A National Threat Map**	Static bases are being established in Poland and the EU. The data include the number and frequency of terrorist crimes and attacks. Threat maps are created. In Poland, **a National Threat Map** is created. The data for the National Threat Map is entered by the citizens themselves. In Poland, data analysis using safety maps is used. These are spatial data. In Poland, a system is being developed to cover the statistics of crimes and terrorist incidents. However, these are only statistical data, which are partly used to construct risk maps and threat maps.

20	**TMNL**	which is a private initiative between banks – local initiative: https://tmnl.nl/en/
21	**SCSN**	Outside the FCT domain: part of the International Data Spaces initiative, in which TNO is an active partner. Scottish Community Safety Network – The home of the Scottish Community Safety Network (safercommunitiesscotland.org)
22	**'I-spaces' in the EUHubs4Data project**	BDVA/DAIRO grants a label for European Data Innovation Spaces and Hubs to help drive forward Big Data adoption and AI based innovation across all domains within European industry. Every year, existing hubs can obtain a label as a "BDVA/DAIRO i-Space", after an evaluation process based on a comprehensive criterion catalogue. https://www.bdva.eu/labelled-i-spaces-2021
23	**The NL AI Coalition**	Has a Data Sharing working group. They focus on setting up data spaces in different domains. Algorithms that work for everyone | NL AIC | The Netherlands AI Coalition
24	**The Data Sharing Coalition**	Has started in NL and is now expanding internationally. This group focuses on providing practical tools and approaches and sharing best practices. Home - Data Sharing Coalition
25	**Schengen information system**	The Schengen Information System (SIS) is the most widely used and largest information sharing system for security and border management in Europe. As there are no internal borders between Schengen countries in Europe, SIS compensates for border controls and is the most successful cooperation tool for border, immigration, police, customs and judicial authorities in the EU and the Schengen associated countries. Schengen Information System (europa.eu)
26	**INTERPOL – Unified Information Architecture**	The Unified Information Architecture project will break down information siloes by creating a common information repository for data. This centralization will allow data to be interlinked and therefore provide improved intelligence to law enforcement. Access management will also be consolidated to ensure data protection and security. The project will benefit all National Central Bureaus and authorized law enforcement agencies in all INTERPOL member countries. Unified Information Architecture (interpol.int)
27	**DEP Data Spaces**	(https://www.bdva.eu/node/1906)
28	**International Data Spaces**	There are many data space calls for various domains. Cloud-to-Edge infrastructure and services seems relevant. (https://internationaldataspaces.org/)

Table 4: List of projects relevant to LAGO mentioned by experts.

	Project/Initiative name	Description	Website
1	**AP4AI (Accountability Principles for Artificial Intelligence)**	The AP4AI Project addresses this challenge by creating a global Framework for AI Accountability for Policing, Security and Justice.	www.ap4ai.eu
2	**AIDA**	AIDA is a Horizon 2020 funded EU project which aims to deliver a descriptive and predictive data analytics	https://www.project-aida.eu/

		platform and the related tools to prevent, detect, analyse, and combat cybercrime and terrorism.	
3	**Anti-FinTer**	Anti-FinTer aims to improve law enforcement capabilities, increase capacity, and develop expertise in the area of terrorist financing associated with activities in dark web, crypto-assets, new payment systems and darknet marketplaces.	https://anti-finter.eu/
4	**APPRAISE**	APPRAISE is a Horizon 2020 funded EU project which aims to create a robust security framework for cyber and physical protection of public spaces, and builds on the latest advances in big data analyses, AI, and advanced visualisation	https://appraise-h2020.eu/
5	**CC-DRIVER**	The CC-DRIVER project seeks to understand the drivers of cybercriminal and research methods to prevent, investigate and mitigate cybercriminal behaviour.	https://www.ccdriver-h2020.com/
6	**COPKIT**	The COPKIT project focuses on the problem of analysing, investigating, mitigating, and preventing the use of new information and communication technologies by organised crime and terrorist groups. For this purpose, COPKIT proposes an **intelligence-led Early Warning (EW) / Early Action (EA) system for both strategic and operational levels**.	https://copkit.eu/
7	**CYBERSPACE**	This project aims to provide policymakers, law enforcement agencies and the private sector with a more comprehensive understanding of cyberattacks and cybercrime in the EU. Insights will be used to develop investigative tools, improve information sharing, and better detection, response, and prevention of cybercrime.	https://cyberspaceproject.eu/
8	**CYCLOPES**	Cyclopes aims to develop to build a network, monitor innovation and research and search for products and solutions useful to combat cybercrime.	https://www.cyclopes-project.eu/
9	**CumuluZ**	University medical centres are developing a new target architecture for Dutch healthcare data with CumuluZ. This project aims to unlock data in a platform consisting of regional hubs, which will facilitate data-driven care, artificial intelligence, network care, prevention, and derived use for scientific research.	https://www.cumuluz.org/
10	**CMINE (Crisis Management Innovation Network Europe)**	CMINE is a hub for crisis management professionals in the EU and beyond. It aims to foster innovation and research uptake in crisis management through cross sector, multi stakeholder dialogues around capability gaps and potential solutions.	https://www.cmine.eu/
11	**Cut The Cord CTC**	Cut The Cord (CTC) project aims to prevent and predict, while assisting Law Enforcement Agencies and other entities to fight financial crimes and "cut the cords" to non-	https://ctc-project.eu/

		traditional products for financing and supporting terrorist organizations.	
12	**European Finance Data Space**		Register of Commission expert groups and other similar entities (europa.eu)
13	**European Health Data Space**	To unleash the full potential, od health data, the European Commission is proposing a regulation to set up a data space that: supports individuals to take control of their own health data, supports the use of health data for better healthcare delivery, better research, innovation and policy making and finally, enables the EU to make full use of the potential offered by a safe and secure exchange, use and reuse of health data.	https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en
14	**Eurojust**	Eurojust works with national authorities to combat a wide range of serious and complex cross-border crimes involving two or more countries. The Agency leads the judicial response to growing threats in Europe, mainly focusing on organised crime groups.	https://www.eurojust.europa.eu/
15	**European Open Science Cloud (EOSC)**	The European Open Science Cloud (EOSC) aims to provide European researchers, innovators, companies, and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation, and educational purposes.	https://eosc-portal.eu/about/eosc
16	**EUROPOL**	Europol attends to EU Member States needs and analyses crime trends in the EU. The Agency supports investigations initiated by Member States. Its work consists of dealing with crimes that require an international approach and cooperation between several countries, inside and outside the EU. The decision on which crimes to prioritise is shaped by EMPACT.	https://www.europol.europa.eu/about-europol
17	**EUROPOL Platform of Experts**	The Europol Platform for Experts (EPE) facilitates the sharing of best practices, documentation, innovation, knowledge, non-personal data on crime.	Europol Platform for Experts (EPE) | Europol (europa.eu)
18	**FREETOOL**	The FREETOOL Project brings together developers and offers them a platform for collaboration and distribution and developing free tools for law enforcement personnel.	FREETOOL — Free Reliable Tools For Investigating Cybercrime (thefreetoolproject.eu)
19	**FRONTEX**	Frontex, the European Border and Coast Guard Agency, supports EU Member States and Schengen-associated countries in the management of the EU's external borders and the fight against cross-border crime.	https://frontex.europa.eu/about-frontex/who-we-are/tasks-mission/
20	**GAIA-X**	GAIA-X is a European project, bringing together research institutions, business institutions, administrations, and politics in the creational of an	https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

		open, transparent, and secure data infrastructure and digital ecosystem.	
21	**Health RI**	Health-RI is a public-private partnership of organizations involved in health research and care. Its mission is to create better health for citizens and patients by reusing health data with an integrated health data infrastructure for research and innovation.	https://www.health-ri.nl/
22	**INFINITECH**	INFINITECH is a joint effort of global leaders in ICT and finance towards lowering the barriers for BigData/IoT/AI driven innovation, boosting regulatory compliance, and stimulating additional investment	Infinitech - The Flagship Project for Digital Finance in Europe (infinitech-h2020.eu)
23	**LOCARD**	LOCARD aims to develop a holistic platform aimed at ensuring the chain of custody throughout the flow of forensic analysis. It is a distributed and trusted platform that allows the storage of digital evidence metadata using blockchain	https://apwg.eu/locard/
24	**OCRE (Open Clouds for Research Environment)**	OCRE is a Horizon 2020 funded EU project, aiming to accelerate cloud adoption in the EU research community.	https://www.ocre-project.eu/
25	**RAYUELA**	RAYUELA project was created to empower and educate young people (children and teenagers primarily) in the benefits, risks and threats linked to the use of the Internet by playing, thus preventing, and mitigating cybercriminal behaviour.	https://www.rayuela-h2020.eu/
26	**SCSN**	Outside the FCT domain: part of the International Data Spaces initiative, in which TNO is an active partner	
27	**STARLIGHT**	STARLIGHT is a Horizon 2020 funded EU project, aiming to create a community that brings together LEAs, industry, researchers, and practitioners in the security ecosystem and to bring AI into operational practices. H2020 - project pays considerable attention to the collection and generation of datasets that are relevant for the development and evaluation of Artificial Intelligence tools. Currently, the project has created some ~30 datasets and identified ~100 existing and/or public datasets that are useful for research in the FCT domain.	https://www.starlight-h2020.eu/
28	**Tech Against Terrorism**	Tech Against Terrorism is an initiative supported by the United Nations, aiming to support the tech industry in building capacity to tackle the use of the internet for terrorist purposes whilst respecting human rights.	https://www.techagainstterrorism.org/project-background/
29	**The European Interoperability Framework**	The National Interoperability Framework Observatory (NIFO) is one of the mechanisms put in place by the European Commission as a monitoring tool, to regularly gather information on the state of	https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-

		play of digital public administration and interoperability activities within the EU Member States and associated countries.	interoperability-framework-detail
30	**Tools4LEAs**	The Tools4LEAs project aims at establishing a long-term and sustainable structure that delivers on a regular basis tool that are ready to be used at operational level by European public security practitioners (law enforcement agencies, forensic institutes, and others), primarily in their digital investigations. These tools will have no license costs for European public security practitioners.	https://www.eactda.eu/projects/Tools4LEAs/home.html
31	**TRACE**	TRACE solutions enable Law Enforcement Agencies to detect and combat money-laundering operations and financing of organised crime and terrorism by increasing the efficiency of information sharing.	https://trace-illicit-money-flows.eu/

Table 5: Relevant practices, strategies, and regulations mentioned by experts

Initiative / Name	**Domain**	**Type (Regulation, Practices, Strategy, Initiative, etc.)**	**Level (EU, National, Organisational, etc.)**	**Overview**
Data exchange between Europol and Private Parties	FCT & Forensics	Practices	EU	Current practices of data exchange between private parties, Europol, and national LEAs and their gaps.
The Prüm Convention	FCT & Forensics	Law	EU	Member states can share information, such as biometric data, with other Member states when investigating crime, terrorism, or illegal migration.
MOD Data Strategy for Defence	Defence	Strategy	UK	Strategy describing the future improvements in Defence Data, including technological solutions and personnel trainings.
Scottish Safe Havens	Healthcare	Infrastructure	Scotland	Trusted research environment for Scottish Healthcare which feeds in the NHS TREs.
NHS TREs	Healthcare	Infrastructure	UK	Provides approved researchers form trusted organisations with timely access to health and care data.

4 Relevant Regulations and Standards

Section 4 provides an overview of relevant regulations and standards that partially must, and partially could inform, the architecture of the LAGO RDE. This section does not aim to be a comprehensive description of regulations and standards but provide pointers and concise summary to indicate regulations and standards emerging from the document review by LAGO partners and expert consultations. The summaries rely on and were partly directly supplied by LAGO partners in this task, albeit shortened were appropriate. The overview considers regulations and standards on EU level, as well as national level (see also section 2 on document analysis methodology).

4.1 EU Level

According to the General Data Protection Regulation (GDPR), personal data shall be "*kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1)*[1] *subject to implementation of appropriate technical and organisational measures required by the Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation')*" (The European Parliament, 2016). Additionally, the Data Protection Act (DPA) 2018 puts control over how personal data is used by organisations, business, or the government (GOV.UK, n.d.). The DPA has rules that must be adhered to which ensures that information follows strict principles such as: *information is used fairly, lawfully, and transparently, information must be accurate and, where necessary, kept up to date* and *information is kept for no longer than necessary.* The DPA also has put stronger legislative protection on sensitive information such as *biometrics (when used for identification), race, ethnic background, health,* and *genetics.* The DPA gives individuals the right to find out what personal information the government and other organisations have stored.

4.1.1 Personal Data Exchange Practices between Europol and Private Parties

The consulted experts indicated of a report concerned with the data exchange practices between Europol and private parties. As of the Europol Regulation, Europol is prohibited from transferring personal data directly to private parties, except in three cases, one of them being the "system of referrals". Europol is also allowed to transfer publicly available personal data to private parties, and to transfer non-publicly available personal data if necessary for preventing and combating internet-facilitated crimes (Milieu Consulting, 2020).

The Europol Internet Referral Unit (EU IRU) is in charge for flagging online terrorist content for referrals and send them to Online Service Providers (OSPs). The OSPs are not legally obliged to take down online content (although they often do), and they check the content against their own terms of reference (Milieu Consulting, 2020). Upon submission of the referral, EU IRU often received an automated response from the OSP stating its safe receival. More in-depth responses including personal data might be received at a later stage. The system of referrals aided the removal of substantial amounts of terrorist content from the online

[1] Article 89 GDPR https://gdpr-info.eu/art-89-gdpr/

space and established a good working relationship between OSPs and Europol, which might encourage OSPs to share more relevant investigative data beyond the one included in the referrals (Milieu Consulting, 2020). However, the process is currently hindered by the current regulations which state that Europol can process the referred data only for the sole reason of identifying the relevant Europol National Units (ENU), which might be a challenging process if the data shared by the OSP is not specific enough. Europol also is not legally allowed to directly seek more information from third parties in order to identify the ENU. Additionally, if Europol fails to identify the relevant ENU within four months of receiving the data, the data needs to be deleted, leading to missed opportunities. Upon the successful identification of the responsible ENUs, the data may not reach again Europol as there is no legal obligation for ENUs to resubmit data once it reaches them (Milieu Consulting, 2020). A revision of the Europol Regulation regarding sharing personal data should be implemented, and Europol should be allowed to exchange data with OSPs for analysis purposes. Such changes will require increased capacity on Europol’s end which would enable the sufficient management of the extended role (Milieu Consulting, 2020).

Usually, Europol receives data from third parties through intermediaries. Third parties are legally obliged to share personal data of investigative value with the national LEAs, which on their end might transfer the data to the Europol National Units (ENU) and the information may be further transferred to Europol itself. However, only a small amount of the personal data shared with national LEAs reaches Europol (Milieu Consulting, 2020). As identified by Milieu Consulting, 2020, the main reason for that is the lack of competence on the side of LEAs or ENUs to act on the case in question, which might be derived from a lack of legal basis to start an investigation. Other issues that hinder personal data from third parties to reach Europol include the insufficient legal basis for third parties to transfer personal data to national LEAs and the legal framework preventing third parties to share all personal data on multi-jurisdictional level with all LEAs concerned (Milieu Consulting, 2020). Thus, some of the recommended solutions include legally enabling Europol to exchange personal data directly with third parties and the amendment of the current EU regulations to allow private parties to share personal data with LEAs on more grounds, along with the integration of a platform that would enable data exchange between stakeholders within the same domain. The implementation of the recommended solutions would decrease the missed opportunities currently faced due to the slow process.

Another method of personal data transferring is the proactive sharing where third parties directly send personal data to Europol. However, Europol is obliged to transfer the data to the relevant ENU, which may decide to resubmit the data back to Europol and include the national LEAs in the process (Milieu Consulting, 2020). The system of proactive sharing is rarely used and often, data ends up being transferred back and forth between the LEAs (if involved) and ENUs before reaching Europol again. This results in a slow process, eventually rendering the data irrelevant and leading to missed opportunities. The role of Europol should be extended and allow them to process the received data beyond the sole purpose of identifying the ENUs. These changes would suggest that the capacity of Europol should be increased, and the data protection and safeguard measures should be adjusted to reflect the enhanced data processing activities of Europol. Proactive sharing might also be burdening for OSPs as it often requires them to process the data before submitting it to the relevant jurisdiction, which often have different standards for the received datasets (Milieu Consulting, 2020).

4.1.2 The Prüm Convention

The Prüm convention is an agreement between Belgium, Germany, Spain, France, Luxembourg, the Netherlands, Austria and later Estonia, Finland, Hungary, Romania and Slovenia to enable and improve cross-border collaboration and cooperation in the field of crime, terrorism, and illegal migration (Council of the European Union, 2005). The Prüm Convention will enable countries involved the opportunity to access DNA records, fingerprint information and any other relevant information if needed (Council of the European Union, 2005). The data sharing agreement works on a "hit/no-hit basis" where a Member state can request further information through the appropriate methods, with personal information only being exchanged once a match has been identified (Machado & Granja, 2018).

The proposal of Prüm II further builds on the original framework, enabling data sharing between members to be quicker, as well as, increasing the level of security and privacy when sharing information (European Commission, 2021). The amended Prüm II includes the addition of new categories such as facial imaging records and police records, as well as the automatic exchange of this information (European Commission, 2021). The update and amendment under Prüm II will ensure that relevant data is available for LEAs in one Member state is available for LEAs in any other Member state, as well as guaranteeing that Europol can also aid Member states in the context of the Prüm framework (European Commission, 2021). Prüm II will support the development of "a new architecture that allows for easier and faster exchange of data between Member States and that ensures a high level of protection of fundamental rights" (European Commission, 2021).

4.2 National Level

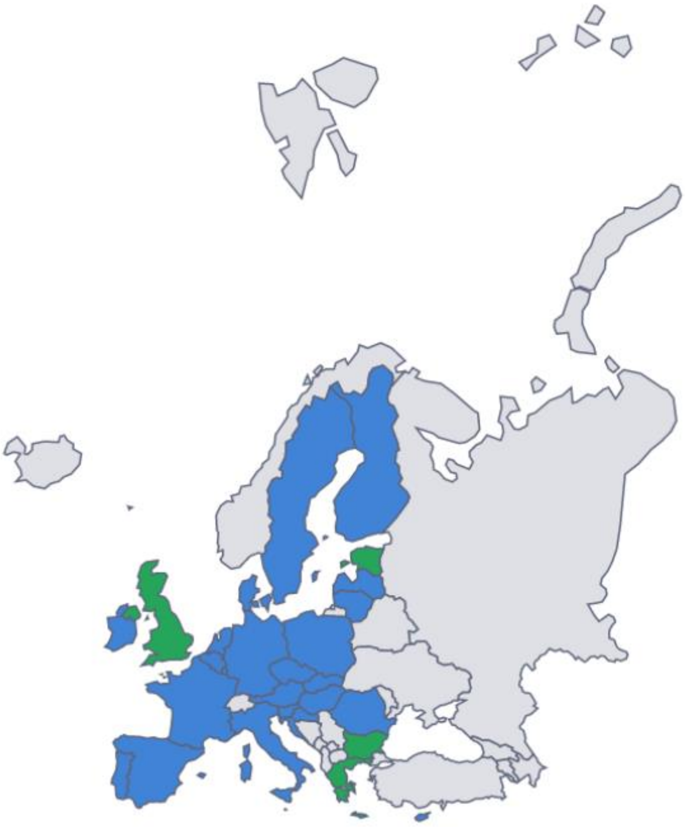


Figure 4: Map of some of the areas covered in this section: EU level (blue and green); national level – UK, Greece, Estonia, Bulgaria (green)

4.2.1 United Kingdom

4.2.1.1 Trusted Research Environments

In the United Kingdom, the Department of Health and Social Care's Data Strategy (UK Department of Health and Social Care, 2022) outlines the utilisation of healthcare data and points to more effective ways to utilise the National Health Service's (NHS) data for research and development. A pivotal point of this process is improving the trust of the citizens in the use of data in the health and care system. The confidence of the data subjects regarding how their data is handled can be increased by ensuring the data is processed in a safe and secure manner, by providing a good collective understanding on how subjects' data is processed and presenting the benefits, both on individual basis and for improving the population health by research and innovation, and by providing a greater access to individuals' personal data and the power to decide how it is used (UK Department of Health and Social Care, 2022). This can be largely achieved by utilising Secure Data Environments, or Trusted Research Environments (TREs). TREs provide a joint system where data is securely stored, accessed, and analysed in-situ, all whilst ensuring high levels of transparency and security. Tres largely adhere to the 'Five Safes' model (Desai et al., 2016): Safe people, Safe projects, Safe data, Safe outputs, and Safe settings. Data can be accessed and processed only by users whose credentials have been established by an accredited authority ('Safe people'), and the data do not leave the TRE ('Safe setting) (UK Health Data Research Alliance, 2021). The nature of TREs implies that sensitive and/or personal data is stored and processed in them, and they can vary in size. For instance, the Genomics England Research Environment (*Genomics England*, n.d.) stores more than 20 petabytes of genome data, and SAIL DataBank (*SAIL Databank*, n.d.) stores the data of the Welsh population collected for up to 20 years and including the records of over 5 million people.

The set of processing tools provided within the TRE varies greatly depending on the general needs of the users. Sometimes, the provided tools may not suffice and fulfil the requirements of the researchers, which implies that there should be possible for researchers to introduce their tools and algorithms to the environment whilst preserving the safety of the resources. A safe setting research environment should have the so-called 'Air lock' capability, which provides the researcher with a build-in set of data analysis tools to prevent the uncontrolled import and export of software and data. However, the 'air lock' capability should allow the researchers to import tools and outside data in the safe setting, as scanning and checking of the imported materials is paramount to ensure that they comply with the requirement, do not tamper the security and integrity of the TRE, and do not record or export data (UK Health Data Research Alliance, 2020a). The established by ONS safe setting Virtual Microdata Laboratory (VML) allows researchers to analyse data on the system whilst restricting any access to the internet, email, printers, or any other means to export data. The VML benefits from being located in secure rooms with CCTV and having an integrated monitoring software which records every click of the mouse and keystroke (Stokes, 2017a). DARE UK discusses the application of Virtual Desktop Interface for access of a safe setting where, similarly to the VML, the researcher would not have access to websites that are not specifically whitelisted and opened in a read-only setting (UK Health Data Research Alliance, 2020a).

To ensure Safe data, the data is de-identified or anonymised before access is granted to researchers. The employed processes of anonymisation or de-identification should minimise the risks of re-identification of the individuals (UK Health Data Research Alliance, 2021).

The use of TREs can largely benefit the broad research and healthcare community; thus, improving the quality of life of the general public. There are a number of examples of successfully operating TREs across UK, such as the Charter for Safe Havens in Scotland (The Scottish Government, 2015), OpenSAFELY (*OpenSAFELY*, n.d.), SAIL Databank (*SAIL Databank*, n.d.), and more. TREs facilitate federation and collaboration and provide the required resources for processing, whilst simultaneously strictly controlling what tools and resources from the "outside" can be introduced to the environment and what data and outputs can be extracted from the environment (UK Health Data Research Alliance, 2021). Except for the increased security, TREs also present a number of practical and cost saving benefits by maximising the use of High-Performance Computing whilst decreasing the costs involved with data transferring and storing (UK Health Data Research Alliance, 2020b).

4.2.1.2 DARE UK Landscape

As part of the DARE UK programme (DARE UK, 2021a), a landscape review was performed to summarise the key unmet needs and opportunities within the UK research and innovation ecosystem. Part of this research includes the conduction of 60 interviews with stakeholders from a spectrum of disciplines and their colleagues, with a total number of 79 interviewees. The needs of the researcher and technologist communities, represented by the interviewees, reflect the problems connected with the creation and maintenance of digital research infrastructure and the access to it (UK Research and Innovation et al., 2021). The unmet needs outlined during the interviews are grouped into six themes: 1) data and discoverability; 2) access and accreditation; 3) digital research infrastructure; 4) capability and capacity; 5) demonstrating trustworthiness; and 6) funding and incentives.

4.2.1.2.1 *Data and discoverability*

The first recorded unmet need is concerned with the data standards and its discoverability. Although the use of technical standards can lead to interoperability of data, there are multiple sets of standards concerning data in the UK alone, each of which used by a limited number of parties, making data not interoperable (UK Research and Innovation et al., 2021). The committed collaborations of bodies (e.g. universities and hospitals) can be an important aspect in the development, enhancement, and implementation of more widely used data standards, leading to more interoperable data.

Another aspect that can reduce the usefulness of data is the poor recording and missingness of its features. Datasets may remain under-utilised if they are not well-documented, making them less discoverable. Vice versa, well-documented datasets can be available via multiple TREs which may lead to increased costs (UK Research and Innovation et al., 2021). The development of user-friendly sets of metadata to describe datasets can make data more discoverable, which can be further enhanced by implementing infrastructures that allow metadata sharing and services for browsing different types of data.

4.2.1.2.2 *Access and accreditation*

Inconsistent or missing standard accreditation processes lead to slowed down data access, meaning slowed down research and innovation and greater administrative burdens. The interviewees from DARE UK highlighted specific areas where implementing widely applied standards can be beneficial for the research communities, including information security, platform specifications, and a centralised codified approach to data licensing. Strong support is also observed to providing platforms which are more accessible to researchers across disciplines (UK Research and Innovation et al., 2021).

Another issue outlined by the DARE UK interviewees is the lack of clear definition of a TRE; however, the above described Five Safes framework is defined as the basis of multiple UK TREs (Stokes, 2017a). An obstacle that researchers encounter is the necessity to gain multiple accreditations by completing trainings for accessing the TREs, and often times these trainings may be duplicative and leading to time consuming. An opportunity improving the current situation can involve a process with key stakeholders and research councils where clear definitions of TRE and the related processes are agreed on. Additionally, the standardisation of the researcher accreditation in the direction of unilateral recognition can greatly improve the experience of the stakeholders, where the trainings are conveyable across various TREs and the standardised researcher accreditation can be used as a TRE pass (UK Research and Innovation et al., 2021).

4.2.1.2.3 *Digital research infrastructure*

Current physical and software infrastructures often are not interoperable and vary widely depending on data types, user requirements, and subject areas, leading to siloed work between research organisations and disciplines and lack of clarity of the data available outside of the different spheres. Present TREs do not facilitate cross-disciplinary research and the linking of data from multiple fields. On a technical level, TREs have an irregular demand on compute power and lack common requirements, resulting in more delays for researchers and often times limiting them to work in a specific geographic location. A suggested solution of the abovementioned issue is bolting high performance computing (HPC) capability to TREs (UK Research and Innovation et al., 2021).

4.2.1.2.4 Capability and capacity

A reported issue by the interviewees is the increasingly growing skills shortages that institutions face, resulting in the need for researcher trainings across disciplines, especially in the technical aspects of using TREs, such as coding for large scale analyses and good data management skills. Another observed shortage is the one of individuals creating digital research infrastructures, such as data scientists and statisticians, which can be improved by furthering the collaborations between public and private sector (UK Research and Innovation et al., 2021).

4.2.1.2.5 Demonstrating trustworthiness

Trust between the different stakeholders involved in research, such as data subjects, data custodians, researchers and funders, is at the very core of research. However, public concerns are present regarding the potential risks involved with data sharing, especially regarding the commercial access to data. Researchers and data custodians have the opportunity to ensure the public that utmost data security, safe use, and compliance are in place by engaging with the public. The aforementioned issue of staffing capacity increases the issues associated with the responsibilities with data custodians. Data custodians are responsible to safely manage the data access and often deal with a greater volume of data access request than their designated capacity. This may lead the research community to believe that risk management is not up to par. Additional delays may be experienced by researchers due to the lack of standardised risk management frameworks. Amongst the proposed opportunities by the interviewees of DARE UK to meet the needs of the research community is the improving of the efficiency of data access requests processing by using the services of entrusted legal teams, governance teams, and contracts teams. Another proposal suggests the application of platforms for managing and prioritisation of time-sensitive applications (UK Research and Innovation et al., 2021).

4.2.1.2.6 Funding and incentives

Standard research funding time frames are limited, especially compared to data access processes and reviews, which often results in researchers fitting within the available and accessible data. Another challenge in data sharing is the lack of clarity regarding the rights and responsibilities of the contributing organisations, often making it unclear which body is responsible for data quality. An opportunity to tackle this issue is by giving credit and recognition (such as co-authorship and acknowledgements) to the organisations which invest in structured data collection and those who maintain and curate data (UK Research and Innovation et al., 2021).

4.2.1.3 The SafePod Network

Administrative Data Research UK (ADR) is a partnership between ADR England, ADR Northern Ireland, ADR Scotland and ADR Wales and the Office for National Statistics (ONS) (ADR UK, n.d.-a). ADR UK has developed a 'SafePod Network' (SPN) which enables researchers to securely access the UK's public sector data for research which will benefit the public (ADR UK, n.d.-b). The creation of the SPN allows researchers to overcome time and cost barriers that may come alongside accessing data securely. The network also enables data centres to securely share their data on a wider scale. Once an individual has their research approved by a data centre they can securely access project data from a SafePod, with each data centre setting regulations for the access of a SafePod as well as the data stored in the SafePod. By March 2023 there will be 25 accessible SafePods that researchers can utilise, with more planned to be installed, the SafePods are primarily located at universities across the UK (SafePod, n.d.). The SafePod Network has security and governance requirements that it must meet which are aligned with each Data Centre, the SafePod Network also maintains the IT systems and relevant equipment, which ensures that the SafePods are kept up-to-date and remain secure.

4.2.1.4 Scottish Safe Havens

Across the UK, and globally, Trusted Research Environments (TREs), or ‘Safe Havens’ exist as a secure environment where approved researchers can access sensitive data and datasets to conduct pre-approved research projects. A TRE provides a trusted environment where researchers are expected to ensure the protection of personal data and the protection of the TRE. Individuals with access to the TRE are expected to adhere to the regulations and principles set by the organisation running the TRE (Kavianpour et al., 2021a). Most TREs use a private cloud infrastructure and also use on-site servers which simplifies any data regulations.

In Scotland, Local Safe Havens operate in the Scottish regions of Aberdeen, Dundee, Edinburgh and Glasgow. The Local Safe Havens feed into the NHS’ TREs which provides a secure environment where health data can be linked to other health data which then provides opportunity for analysis and research to be conducted (DARE UK, 2021a). The Scottish Government has created a ‘Charter for Safe Havens in Scotland’ which covers the infrastructure of the Safe Haven to ensure that it remains compliant with its regulations as well as relevant legislation (The Scottish Government, 2015).

Aberdeen – The Grampian Data Safe Haven (DaSH)

DaSH started in 2012 by the University of Aberdeen and NHS Grampian and provides a secure setting for linking data and hosting data by accessing through a VPN. DaSH is accredited by the Scottish Government and meets the Information Security and Governance regulations stated in the Charter for Safe Havens.

Dundee – Health Informatics Centre (HIC)

The HIC has been labelled as “a leader in health data linkage” (DARE UK, 2021). The HIC manages a repository of eHealth data which covers approximately 20% of the Scottish population. The HIC enables remote access to the Safe Haven where users are required to access the data through a server at HIC where the user is given secure and remote access to conduct analysis.

Edinburgh – DataLoch

DataLoch routinely collects data from individuals during their daily interactions with the health and social care services. Currently DataLoch is enabling academics and health and social care professionals to submit applications to enable them to access data.

Glasgow Safe Haven

The Safe Haven allows researchers to access anonymous health datasets, a data analytics platform and provides expert support to enable "data drive discovery with de-identified NHS data" (DARE UK, 2021a). The Glasgow Safe Haven is accessed through a VPN or a safe room at the University of Glasgow.

4.2.1.5 Defence

United Kingdom's Ministry of Defence (MOD) is working towards improving the national defence data strategies. The character of warfare changes due to the cross-domain digitalisation, which inevitably affects the way Defence data is collected, processed, stored, and shared (UK Ministry of Defence, 2021d). Presently, the UK MOD faces multiple data challenges, including:

- Inaccessible data in internal/contractual silos;
- Lack of full recognition on the importance of data;
- Digital and data knowledge and skills gaps;
- Lack of standardised data exploitation and delivery;
- Inconsistent governance and control;
- Fragmented, insecure, and fragile technology core (UK Ministry of Defence, 2021d, 2021b).

The MOD's *Data Strategy for Defence* (UK Ministry of Defence, 2021b), *Digital Strategy for Defence* (UK Ministry of Defence, 2021d) and *Defence Data Management Strategy* (Ministry of Defence, 2020b) apply to the Defence and wider data ecosystem, including international allies (i.e. NATO), industry, agencies, and academia. All the aforementioned strategies are ongoing, with prospective outcomes by 2025 (UK Ministry of Defence, 2021b) and 2030 (UK Ministry of Defence, 2021d). The purposes of data sharing in Defence are mainly for exploitation and data are used as strategic assets, as well as for research and innovations in the domain.

Defence uses a variety of data, be that structured, unstructured, semi-structured, or images. The strategies envision that data from the battle and business space, including digital data and sensor data, can be brought together by the Digital Backbone (UK Ministry of Defence, 2021d). Furthermore, Defence data can be divided to Core and Non-Core Defence Data, as the main difference between the two types is that non-Core data is not shared outside its originating area, whereas Core data supports critical Defence processes and can be shared across multiple levels (Ministry of Defence, 2020b).

Currently, Defence data is hard to access and integrate because it is fixed inside internal and contractual silos (UK Ministry of Defence, 2021d, 2021b). The digital strategy states the building and delivery of a single Digital Backbone for Defence, which includes a Hyperscale Cloud, Next Generation Networks, and User services, which will allow the right people to access the right data from the battlefield, office, or headquarters, from all Defence domains (Marine, Air, Land, Space, and Cyber) (UK Ministry of Defence, 2021d). The Hyperscale Cloud will allow the advanced applications and services required by Defence in a timely manner, whilst enabling personnel to access and process data securely and quickly on the battlefield and allow the business space to run systems directly on the Cloud. Generally, such cloud architecture will enable data storage, processing, and sharing in a rapid manner whilst allowing users to access the cloud, regardless of their whereabouts (Ministry of Defence, 2020b). Additionally, the Hyperscale Cloud will be able to provide the foundations for the development, adoption, and enhancement of innovative technology as it is envisioned to be able to operate at multiple classification levels (UK Ministry of Defence, 2021d). The strategies state that security will be an integral part of the Digital Backbone by adhering to new architecture standards and governance processes; however, there are no other specifics mentioned (UK Ministry of Defence, 2021d).

As abovementioned, data and digital literacy across Defence need to be enhanced. Thus, all three strategies propose enhanced support, education and training of all grade and rank personnel on data, their analyses, and the technologies driven by data (Ministry of Defence, 2020b; UK Ministry of Defence, 2021b, 2021d). Amongst the strategic outcomes aimed to be achieved by 2030, the delivery of a transparent data portfolio, the establishment of common standards, and the cataloguing and increased accessibility and usability of data are the prime priorities (UK Ministry of Defence, 2021d).

Defence must treat data in accordance with the Data Ruleset, which states that:

1. *Exercise **sovereignty** over data, including accountability and ownership;*
2. ***Standardise** data across the Defence landscape;*
3. ***Exploit** data at the most effective and relevant point in the value chain;*
4. ***Secure** digital data at creation, curation, when handling, storing and transmitting;*
5. ***Curate** data, ensuring it is assured, discoverable and interoperable;*
6. ***Endure** data as an asset beyond individual projects* (UK Ministry of Defence, 2021b).

Some of the abovementioned practices might be specifically beneficial for research data collaborations and the development of an FCT RDE, such as establishing common standards, utilising common architectures, integrating security in the designs of the developing systems, granting user access based on their specific needs and status, and providing relevant and continuous trainings to the users.

4.2.1.6 Consulted experts from the UK

All the UK experts listed multiple initiatives and projects for the creation of data spaces relevant for FCT RDE. *12.5%* of the UK experts mentioned that FCT RDE may not be particularly effective in the context of face recognition due to the significant privacy concerns raised by the technology. All the experts list multiple standards and policies which are deemed applicable by them in the context of regulations of data practices. The implementation of common standards for data creation and types is listed as a top-five enabler in FCT data sharing by *87.5%* of the experts. Additionally, *87.5%* of the experts express that LEAs show reluctancy when sharing data due to legal and ethical considerations.

4.2.2 Greece

The consulted experts indicated of the Hellenic Data Protection Authority's *Processing of Health Data* (Hellenic Data Protection Authority, 2022) contains the current legislation (directives, laws, etc.) for the processing of personal healthcare data and the limitations involved with the processes. The document is applicable to health services and stakeholders at a national (Greece) level. Due to the nature of personal health data, the sensitivity of the discussed information is high. The directive mentions, without specifying, that multiple procedures are governed and in-line with the Greek legislation for health data sharing, including the procedures involved with data sharing, methods to ensure privacy and increase security, and the rights of the subjects to consent or object the processing of their personal health data (Hellenic Data Protection Authority, 2022).

Another recommended resource by the experts is the Greek Ministry of Administrative Reconstruction's *4th National Action Plan on Open Government 2019-2021* (Open Government Partnership & Hellenic Republic Ministry of Administrative Reconstruction, 2022) has both a national and multinational scope, and it is applicable to LEAs, ministries, and national bodies. The objective of data sharing, as of the action plan, is to enable open access to data of the Ministry of National Defence and the Ministry of the Interior (along with a number of other ministries and national centres). It aims to coordinate the implementation of the open government commitments through the Transparency and open Government Department, which is responsible for the monitoring and coordination of the participation of Greece in the Open Government Partnership (Open Government Partnership & Hellenic Republic Ministry of Administrative Reconstruction, 2022).

4.2.2.1 Consulted experts from Greece

Out of all the consulted experts from Greece, *33.3%* of them indicated that they are not aware of initiatives relevant for the creation of RDE in the FCT domain, with the opportunity to name examples from other domains with comparable complexity. *16.7%* of the consulted Greek experts were also not aware of Data Strategies relevant for the FCT domain. The majority (*66.7%*) of the consulted experts expressed that an RDE would not be particularly ineffective for any FCT domains. *Half* of the consulted experts from Greece were aware of standards and policies which regulate data practices and are considered good by the experts. All the experts listed at least three types of data they consider relevant for FCT RDE. *50%* of the experts also listed lack of trust as a top-five barrier in FCT data sharing. Standardisation is listed as a top-five enabler in FCT data sharing by *66.7%* of the experts. All of the consulted experts express that trust between partners in data sharing and co-creation in FCT research is an extremely important factor. Half of the experts report that data misuse is amongst the main risks in FCT data co-creation and sharing.

4.2.3 Estonia

4.2.3.1 Big Data in Estonia

The consulted experts indicated the Estonian Academy of Security Sciences' "*Suurandmed: olemus ja kasutamise kitsaskohad*" (Puusalu, 2020) ("*Big Data: Nature and Bottlenecks of Use*"), which is concerned with the current Big Data picture in Estonia. Some of the data types described by Puusalu (Puusalu, 2020) include Big Data, which is unstructured and of vast amounts, digital data, personal data, automatic data, such as footage from police security cameras, and data that can be linked to a specific subject, such as speed cameras footage. The data may be structured or unstructured, thus, the formats are varied – textual data, audio data, footage, statistics, and numerical data, and some of their applications include utilising them for trainings and tests. Various levels of sensitivity are linked with the involved data, including publicly available data and personal data (Puusalu, 2020).

The outlined purpose of data sharing in this document is to enable better access to governmental services for the citizens and decrease discrimination by digitalisation. The size of datasets in question is unspecified, but it is known it is large as the ID cards of the Estonian citizens (1,345,724 as of 2020) are part of it.

Data is obtained by a number of different methods: certain data may be voluntarily shared by the subjects via social media and smart phone apps when they register or use online services (i.e. by agreeing the particular Terms and Conditions), such as online banking; another method is utilising various national databases, such as the residency register; and data can also be obtained from the private sector, however, the data collected by the private sector can be obtained following a mutual agreement or if there is a legal basis for the sharing (Puusalu, 2020). Considering the vast variety of data sources, Puusalu (Puusalu, 2020) mentions a broad list of tools used for sharing data, including free web apps and social media networks and national registers. Often, additional programmes are utilised, and analyses are performed to combine the existing data.

Generally, for the majority of the Estonian governmental systems, X-Road (*X-Road*, n.d.) is used for secure data exchange between organisations, which is in line with the GDPR and the Estonian data regulation policies (Puusalu, 2020).

The document outlines multiple obstacles present. Amongst the first described barriers is that already collected data need to be digitalised before handling. Due to the nature of the data, more specific programmes and systems are required when videos and photos are analysed, and the Estonian organisations tend to use readily available software instead of creating their own, which would increase the security levels and enable application of all required features. Security concerns are also mentioned in the context where private companies collect data from certain devices and networks and are able to use the most of it freely since the subjects had given their consent by agreeing with the Terms and Conditions. Such instances might have detrimental effect on the trust of the public and act as a barrier for sharing and processing data for research (Puusalu, 2020). Some of the means to increase public trust in data sharing are ensuring the citizens that the data is not mishandled by any means, encryption is used, and personal ID-card keys are amongst the security measures in place. Additionally, increasing the public understanding on the topic and providing evidence on how such initiatives may enable economic growth and improve the public life standards may be beneficial (Puusalu, 2020).

In the context of FCT RDE acceptance, the authors outline a number of issues that limit the advancements in this direction, such as limited or lacking regulations, stagnant developments of data spaces, and lack of relevant frameworks. There are difficulties linked with collaborations between organisations within Estonia and an insufficient number of experts in the area, which is further complicated by lack of funds (Puusalu, 2020). A suggestion made in the document to tackle the aforementioned obstacles is the participation of Estonia in EU projects related to data collection, sharing, and processing. Other means for improving the overall acceptance of FCT RDE include establishing relevant field-specific policies and collaborations with relevant external stakeholders from the private sector that can enable the more efficient use of current technologies and the development of new ones.

Bearing in mind the broad spectrum on the study by Puusalu (Puusalu, 2020), mainly generic means of compliance are mentioned, such as the use of Terms and Conditions. Additionally, the Estonian government implements a logging system requiring an identification code where all data processing is recorded along with the name and ID code of the data processor (Puusalu, 2020).

4.2.3.2 Estonian ICT Strategy

Another document recommended by the experts is the Estonian Ministry of Interior's Information and Communications Technology (ICT) Strategy (Estonian Ministry of Interior, 2022), which applies to a broad spectrum of governmental agencies within the Ministry of Interior (MoI), including LEAs, the Recue Board, the Academy of Security Sciences, and Emergency Call Centres. The strategy is already implemented and includes planned activities. The data in this document is applied in a variety of areas within the FCT and security network, including:

- Public space protection;
- Crime prevention;
- Rescue services;
- Emergency calls processing;
- Biometric data applied for personal identification;
- Digital forensics;
- Border protection;
- Criminal investigations;
- Digitalisation of criminal proceedings and
- Interoperability of EU large-scale systems in the domain of Justice and Home Affairs (JHA).

The purposes of data sharing specified in the Estonian ICT Strategy are to improve the services of the MoI organisations and to implement AI in the organisations. The data types mentioned in the document include biometric data, text documents, types of digital forensics, geolocations, CCTV, traffic camera images, Big Data, OSINT (e.g., social media data) (Estonian Ministry of Interior, 2022). The data in this document serves public services support and is mainly operational.

The strategy states that the interoperability of the national systems, the increased digitalisation of criminal proceedings, and the central data storage ensure that data is made accessible to the relevant authorities through their relevant systems (Estonian Ministry of Interior, 2022). Some of the data sharing tools mentioned in the strategy are X-road (*X-Road*, n.d.), the Open Data Portal (*Eesti Avaandmete Teabevärav*, n.d.), and national and EU information systems. Two of the mentioned data processing tools are general AI solutions for video analysis and TEXTA, an automatic annotation tool for text documents and emails and forensic image processing. Additionally, virtual machines are used by digital forensics specialists. Security of the data is ensured by tracking the data and the activities linked with it, and the data storage duration varies between the different databases (Estonian Ministry of Interior, 2022).

The strategy outlines certain barriers which might affect the general research data practices, such as inconsistent data quality, not all data being machine-readable, and complications posed by the GDPR and the draft AI Act. Furthermore, in the context of FCT RDE barriers, the strategy specifies that there are regulatory limitations on data sharing with third parties present. The enablers in this document, both to FCT RDE development and research data practices, are the interoperability of the systems, progressively growing amounts of machine-readable data, and common standards (Estonian Ministry of Interior, 2022).

The procedures for risk assessment and management and compliance follow the ISO 27001 standard for Information Security Management (*ISO 27001 Information Security Management*, n.d.) and its Estonian equivalent. Additionally, the strategy implements a central data access management for all organisations within the MoI along with a zero-trust policy, which assumes that the network trying to access the data is hostile and requires verification (Estonian Ministry of Interior, 2022).

4.2.3.3 The joint action plan for the digitalization of criminal proceeding by the Estonian Ministry of Interior and Ministry of Justice

The Estonian *Joint action plan for the digitalisation of criminal proceedings in the administration of the Ministry of Justice and the Ministry of the Interior* (Estonian Ministry of the Interior & Estonian Ministry of Justice, 2021) strategy (recommended by the experts) is concerned with the data practices of the two ministries. The strategy encompasses current and future practices aiming to improve the currently overall poor digital data exchange and increase the speed and efficiency of the criminal proceedings in Estonia.

Some of the data types mentioned in this joint strategy includes biometric data, digital forensics data, procedure details, and metadata; thus, the sensitivity of the data is noted as “Private”. The data in question can be both structured and unstructured. Some of the formats include textual data, numerical data, photographs, videos, and statistics, and it functions as Live data for criminal proceedings (Estonian Ministry of the Interior & Estonian Ministry of Justice, 2021).

The objective of this strategy is to enable the collection of data from different systems in order to provide a more efficient decision-making in criminal proceedings, all whilst going paper-free. The tools and principles mentioned in the strategy are X-road (*X-Road*, n.d.) (for data sharing), AI, machine learning, e-discovery, and mixing services (for data processing) (Estonian Ministry of the Interior & Estonian Ministry of Justice, 2021). Security and privacy are ensured by adhering to the GDPR and relevant national regulations and legislations.

The strategy outlines two main barriers for research data practices and FCT RDE applications, the first one being certain data access restrictions, and the second one – insufficient amounts of data for training and testing AI models. Technological solutions, on the other hand, can be applied to increase the transparency of data processing, which would act as an enabler to improve the acceptance of FCT RDE and general research data sharing (Estonian Ministry of the Interior & Estonian Ministry of Justice, 2021).

4.2.3.4 Consulted experts from Estonia

All the consulted Estonian experts listed initiatives for the creation of data spaces relevant for FCT RDE. Additionally, *75%* of the experts listed data strategies relevant from the FCT domain. All the consulted Estonian experts listed at least one standard or policy which regulates data practices and they consider good for FCT RDE. The experts listed between one and six types of data they consider relevant for FCT RDE. Half of the consulted experts express that exclusion of specific groups and organisations from the FCT RDE would not be necessary but suggest different levels of access and involvement. All the consulted experts list secure data sharing as a barrier in FCT data sharing. Half of the experts say that LEAs are usually reluctant when sharing data. *75%* of the experts also list the lack of common standards in the context of data sharing and processing as a main risk in FCT research.

4.2.4 Bulgaria

The Bulgarian Policy of Personal Data Protection of the Ministry of Defence and the Bulgarian Army (Karakachanov, 2018a) outlines the rights of the Bulgarian Ministry of Defence administration to process personal data. The policy specifies that data may be shared for national defence and security reasons, to ensure the performing of contracted work duties, to support the Information Policy of the Ministry of Defence, and to provide safety and keep order within the Ministry of Defence and the Bulgarian Army.

There are four categories of personal data specified in the policy:

i. Category 1: includes genetical data, data concerning the race and ethnicity of the individual, their political and religious views, medical health, sexuality, and intimate relations.
ii. Category 2: includes any personal data which may enable the identification of the subject, excluding Category 1 data.
iii. Category 3: includes all other personal data which may enable the identification of the subject.
iv. Category 4: includes anonymised and publicly available data.

According to the policy (Karakachanov, 2018a), security is achieved and maintained by:

- hiring trained personnel to handle the personal data;
- integrating internal audits and quality controls;
- conducting risk assessments;
- limiting the access to the data;
- tracking the actions made in the system;
- and providing continuous trainings.

Additionally, the policy states that the duration of data storage depends on the specifics of the case, and it does not provide a time frame. It is also specified that the individual whose data will be handled must be provided with a notice prior to the data processing (Karakachanov, 2018a).

4.2.5 America

4.2.5.1 USA Fusion Centres

In America, the collaboration between law enforcement agencies (LEAs) has become essential as "offenders can easily move from area to area committing a series of crimes" (Pickering & Fox, 2022. Pg. 733). By undertaking this approach crimes do not have to be processed as an isolated crime, and rather as a string of criminal offences. If a region experiences an increase in crimes, a partnership between local, state and federal LEAs to approach a crime collaboratively alongside other relevant work force, for example the creation of fusion centres after the September 11 attacks. Fusion centres enable the exchange of information and data between state, local, tribal and territorial, federal and private sector partners (U.S. Department of Homeland Security, 2022), they have been developed within the U.S.' Department for Homeland Security (DHS). The fusion centres are able to look through FBI databases, drivers' license records, financial information, firearms licenses' and much more (Monahan, 2009). Fusion centres contribute to the Information Sharing Environment (ISE) by receiving criminal information from the government. The fusion centres then analyse the information from the perspective of their local environment, and then disseminate the data to local agencies, this process helps DHS partners and LEAs to identify and react to any threats (U.S. Department of Homeland Security, 2022).

The fusion centres have four Critical Operational Capabilities:

- **Receive**: Ability to receive classified and unclassified information from federal partners.
- **Analyse**: Ability to assess local implications of that threat information through the use of a formal risk assessment process.
- **Disseminate**: Ability to further disseminate that threat information to other state, local, tribal, territorial and private sector entities within their jurisdiction.
- **Gather**: Ability to gather locally generated information, aggregate it, analyse it, and share it with federal partners as appropriate.

Additionally, fusion centres have the protection of privacy, civil rights and civil liberties as a key priority, with the work force required to undertake training on the privacy and civil liberty issues associated with the work that is conducted within fusion centres, under the 9/11 Commission Act[2] (U.S. Department of Justice, n.d.).

4.3 Summary

Table 6: Overview of relevant sources

Reference name	Type of source	Geographical area covered	Relevant for (domain)	Currently used?	Data sensitivity level	Description of data processing procedures (Y/N)
Data Strategy for Defence: Delivering the Defence Data Framework and exploiting the power of data (UK Ministry of Defence, 2021a)	Strategy	UK	Defence	Implementation by 2025	Various	Y
Digital Strategy for Defence: Delivering the Digital Backbone and	Strategy	UK	Defence	Implemented by 2030	Various	Y

[2] The Implementing Recommendations of the 9/11 Commission Act of 2007 (9/11 Commission Act) https://bja.ojp.gov/program/it/privacy-civil-liberties/authorities/statutes/1283

unleashing the power of Defence's data (UK Ministry of Defence, 2021c)						
Defence Data Management Strategy (Ministry of Defence, 2020a)	Strategy	UK	Defence	Ongoing	Various	Y
Trusted research Environments (TRE) A strategy to build trust and meet changing health data science needs (UK Health Data Research Alliance, 2020a)	Strategy	UK	Research	Ongoing	Private data	Y
ICT Strategy	Strategy	Estonia	Governance (Non-public)	Current use and planned activities.	Operational data, sensitive data.	Y
Estonia's Digital Agenda 2030	Strategy	Estonia	Government and private sector.	Future agenda (ongoing).	Personal data, open data, metadata, reuse of data.	Y
Joint action plan for the digitization of criminal proceedings in the administration of the Ministry of Justice and the Ministry of Interior	Strategy	Estonia	Government area of ministries.	Current use and planned activities.	Private data.	Y
Data Research Infrastructure Landscape: A review of the UK data research infrastructure	Review	UK	Research	Current (future recommendations included)	Private data, sensitive data, confidential data	Y

(DARE UK, 2021b)						
Tech Against Terrorism: Transparency Report: Terrorist Content Analytics Platform (Tech Against Terrorism, 2021)	Report	International (UN Nations)	Terrorism	Ongoing	Sensitive data, personal identifiable data	Y
Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act) (European Commission, 2020)	Regulation proposal	EU	Multiple	Ongoing	Personal data	N
Personal data protection (European Parliament, 2022)	Report	EU	Multiple	Current	Personal data	N
Study on the practice of direct exchanges of personal data between Europol and private parties (Milieu Consulting, 2020)	Report	EU	LEAs, security	Current with future recommendations	Personal data	Y
Data Protection & Transparency: Balancing Europol’s operational needs and the	Report	International	LEAs, security	Current	Personal, organisational, and financial data	N

individual’s right to data protection (EUROPOL, 2023)						
40 Terrorism Databases and Data Sets: A New Inventory (Bowie, 2021)	Dataset collection	International	LEAs, security	Current	Datasets not directly available to the general public	N
Strengthening Europol’s Mandate (Cirlig, 2022)	Report	EU, UK	LEAs, Security	Current	Personal data	N
Privacy-preserving cloud computing on sensitive data: A survey of methods, products, and challenges (Domingo-Ferrer et al., 2019)	Academic paper	EU	Various domains	Current	Personal data	Y
Protection Information Management: Framework for Data Sharing in Practice (Protection Information Management, 2018)	Report	International	Data sharing specifically	Current	Personal data	N
Understanding EU counter-terrorism policy (Voronova, 2021)	Report	EU	FCT	Future	Various	N
FAIR Principles (GO FAIR, n.d.-a)	White paper	International	Research	Current	Various	N

5 Relevant Data Practices

Section 5 provides an outline of the data types and data practices identified in documents and expert consultations with relevance for the LAGO RDE. Data practices will be discussed with respect to data types, data processing and data storage.

5.1 Data Types Required For FCT-RDE

Experts identified five core data types the LAGO RDE will be expected to handle. The top-five most frequently mentioned data types are (in descending order): *Text data, Image & Video data, Sensor data, Network data, Audio data* (cp. Table 7). Additionally, Financial data, CCTV, Darkweb data, Biometric data, Communication data, IoT data, GPS and Maps data, Internet data, Criminal data, Sensor data, Qualitative data, Quantitative data, Incident data, Network data, Personal data, Medical data, data from Blogs and Forums, Client information data, Sensitive data, Classified data, Public data, Open data, Relational databases, Survey data, and Intelligence data were listed, although at lower rates. Following that, the experts were asked the rank the top five most important types of data that LAGO RDE needs to accommodate. Figure 5 shows the first positions in the lists of the experts. The five most mentioned data types relevant for LAGO RDE are Video, imagery, CCTV; Operational data; Criminal data; Text data; and Event data.

Table 7: Five core types expected within the LAGO FCT RDE

Data type and priority	Details
1. Text data	▪ Clearweb, ▪ Darkweb, ▪ Social media, ▪ Messaging services, ▪ External and internal sources like criminal reports.
2. Image and video data	Video data from: ▪ Clearweb, ▪ Darkweb, ▪ Telegram, ▪ Surveillance cameras
3. Sensor data	Real-time information about: ▪ Environmental conditions and ▪ Other situational factors.
4. Network data	Valuable insights into: ▪ Communication patterns and ▪ Network activity.
5. Audio data	From: ▪ Phone calls ▪ Radio transmissions ▪ Audio data/recordings that can be analysed using speech recognition.

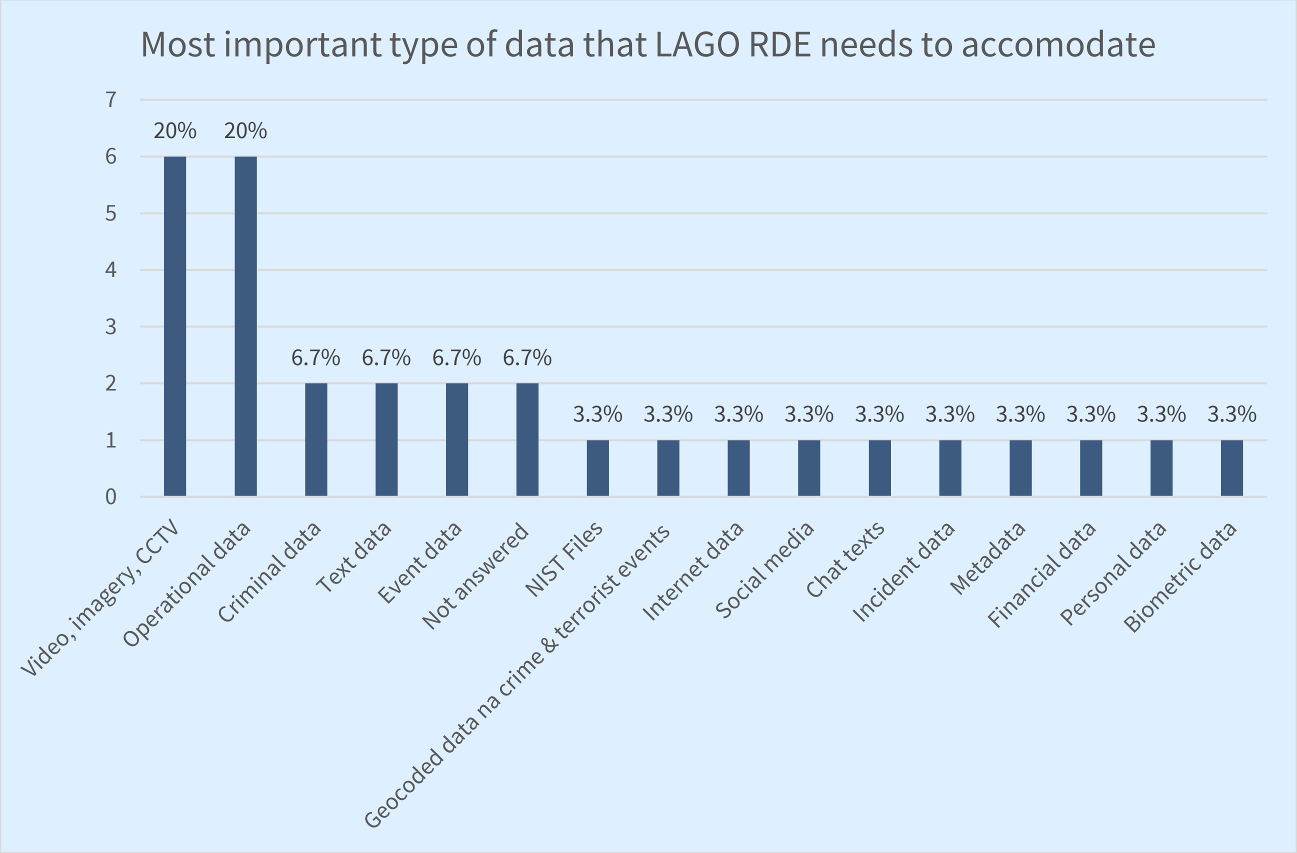


Figure 5: The most important type of data that LAGO RDE needs to accommodate according to the experts.

5.2 Overview of Current Data Practices

5.2.1 Data Processing

Data processing encompasses "*any operations or set of operations which is performed on … data*" (The European Parliament & The European Council, 2016). According to the EU General Data Protection Regulation (GDPR), data processing includes various automated and manual processes, as the recording, collection, organisation, storage, structuring, adaptation, alteration, retrieval, disclosure, dissemination, erasure, alignment, and use of personal data or sets of data (The European Parliament & The European Council, 2016).

The reviewed documents discussed data processing in the sense of research facilitation whilst providing privacy and data protection. The following models and approaches will be discussed in more detail as they seem specifically relevant to the LAGO RDE: The Five Safes, the FAIR Principles, and data protection measures. We include approaches from other areas with similar complexity to FCT research, such as health sciences, as these can provide important lessons.

5.2.1.1 The Five Safes Mode (Health Sciences)

A central challenge in using health data is facilitation of research whilst protecting privacy and managing public trust (UK Health Data Research Alliance, 2020b).The increased emphasis of sharing more sensitive data in health context, resulted by the Covid-19 pandemic, led to multiple investments into research projects concerned with data sharing to support public health decisions and the development of trusted research environments (Tacconelli et al., 2022; UK Research and Innovation et al., 2021). Safe Haven, or Trusted Research Environment, is defined by Kavianpour et al.(Kavianpour et al., 2021b) as 'a secure environment designed for approved, and named researchers to access sensitive data, where access to specific datasets is provided to approved research projects'. For the privacy and confidentiality protection of the data, TRE providers and researchers follow set principles. Examples of such principles are the Health Data Research Alliance Trusted Research Environment Green Paper (UK Health Data Research Alliance, 2020b) and the Charter for Safe Havens in Scotland (The Scottish Government, 2015b).

As a part of the recent/current UK Research and Innovation (UKRI) Digital Research Infrastructure programme DARE UK (the UK Trusted and Connected Data and Analytics Research Environments), the UK Health Data Research Alliance (UKHDRA) research into the approach to data access based around TREs where robust and independent TRE accreditation, auditing, and monitoring are in place (UK Health Data Research Alliance, 2020b). Access to sensitive data is often pivotal for an abundance of researchers. Sensitive data in the context of DARE UK, 'includes personally identifiable information (such as names and addresses), or data which is commercially, legally or politically sensitive or sensitive from an intellectual property perspective' and can also include de-identified data (UK Research and Innovation et al., 2021). The UKHDRA research sets the requirement for basing TREs on the Office for National Statistics' (ONS) Five Safes model (Stokes, 2017b), whilst adding some extensions to the model to reflect the specifics of health data and the latest technological developments. The "Five Safes" in question are: Safe people; Safe projects; Safe settings; Safe outputs; Safe data.

5.2.1.1.1 Safe People

In the context of TREs, Safe people refers to researchers able to demonstrate appropriate credentials who undertake approved (safe) projects (UK Health Data Research Alliance, 2020b). As required by the ONS, safe researchers have to demonstrate that they have the technical skills to use the requested data (either through academic qualifications or relevant experience), they need to complete the ONS training course and pass its assessment successfully. The researchers requesting access to ONS TREs must also agree to their inclusion on a public record and sign a declaration stating that the researchers will protect the confidentiality of the data at all times (gov.uk, 2020; Stokes, 2017a).

5.2.1.1.2 Safe Projects

It is essential to ensure that the use of data is appropriate and has the potential to benefit the public (UK Health Data Research Alliance, 2020b). Thus, trained and accredited researchers need to prepare a research proposal in order to request specific data. In this proposal, the Safe people must ensure the ethical and appropriate use of the requested data, alongside with the delivery of public benefits and published results which would allow further research and use (Stokes, 2017b).

5.2.1.1.3 Safe Setting

There are multiple platforms allowing access to health data through the implementation of safe setting (e.g., Genomics England Research Environment GERE, SAIL DataBank). Safe setting needs, on one side, to promote trust to the public and data controller by implementing security and transparency, whilst simultaneously refraining from burdening the research. A safe setting needs to ensure that individual level data cannot be exported, and that accredited researchers can remotely access the data using their individual accounts whilst the system should be able to keep track of their activity (UK Health Data Research Alliance, 2020b).

5.2.1.1.4 Safe Computing

Since the development of the Five Safes model framework in 2003, many technological developments came in place (Desai et al., 2016). DARE UK outlines Safe computing as an issue that is not explicitly addressed by the Five Safes model but can be seen as an extension of Safe setting. As abovementioned, previously the term Safe setting refers specifically to 'on-premises' hardware where various physical, network, and software security are in place and the system is a responsibility of the TRE operator or data custodian. The described systems can be configured to 'private clouds' by, as aforementioned, allowing researchers to access them remotely by using virtual desktop interface. However, the use of public clouds and third-party computing resources may be a more viable and advantageous option for TREs nowadays. This sort of software and hardware outsourcing may allow short periods of intensive computation and the capacity to engineer scalable platforms (UK Health Data Research Alliance, 2020b). However, as anything concerning sensitive data, ensuring a proficient level of public trust is crucial. This may be accomplished by implementing security designs and engineering along with contractual arrangements with the third-party that would minimise the risk of security breaches.

5.2.1.1.5 Safe Outputs

As abovementioned, a Safe setting implies the presence of 'air lock' preventing export and import of data and software. TREs need to implement systems and processes that ensure Safe outputs by allowing approved data to cross the ‘air lock’ whilst tracking the requests and decisions. Currently, these requests are manually reviewed which could have a negative effect on the research and slow down the process. Thus, the full or partial automation of the reviews of the requested outputs is crucial in the building of a next generation TRE whilst assuring high security levels (UK Health Data Research Alliance, 2020b).

5.2.1.1.6 Safe Return

Another extension of the Five Safes model is the Safe return which may be considered to fall under the Safe setting – Safe outputs category. This subcategory refers to the safe return of research reports to clinical cares and individuals whose data, used in its de-identified form, can be re-identified and increase the convergence of clinical care and research (UK Health Data Research Alliance, 2020b). However, the implementation of Safe return requires return paths where no reports are disclosed to the wrong individuals, thus the necessity of a trusted linkage service is paramount.

5.2.1.1.7 Safe Data

The data accessible to accredited researchers on TREs should be proportionate to the approved project requirements and in line with the General Data Protection Regulations (GDPR). Thus, the potential of identification, be it individual or group, from the data needs to be reduced to minimal (Desai et al., 2016).

5.2.1.2 The FAIR Principles

The FAIR (Findability, Accessibility, Interoperability, and Reusability) Principles act as guidelines for data producers and publishers and give a definition of the constituents of ‘good data management’ and enable data reusability (Wilkinson et al., 2016). The principles apply not only to data but are beneficial to the tools and workflows connected with the data. The application of the FAIR Principles ensures transparency, reusability, and reproducibility. These high-level principles do not suggest any specific standard, technology, or solution, and their main objective is to aid data publishers and stewards in having a more rigorous management of their digital research artefacts by utilising the FAIR principles as a guide (Wilkinson et al., 2016). Often, the processes involved with data adaptation to FAIR is called FAIRification, and there are many aspects of it. The following paragraphs will introduce the different steps that may be utilised in the process of FAIRification.

5.2.1.2.1 Findable

The first step of FAIR – Findable – is concerned with making data findable, both for humans and computers, in order to make it re-usable. These steps are paramount for FAIRification as machine-readable metadata enables the automated discovery of datasets and services.

The four sub-steps linked to Findability are:

F1: *(Meta)data are assigned a globally unique and persistent identifier* (GO FAIR, n.d.-b). This principle refers to the assignment of universally and globally unique persistent identifiers to continuously identify the same resources, even when they no longer exist or are moved. A current challenge with this is ensuring the longevity of the identifier and its persistency even after the termination of the project or community that created the identifier. An example for a globally persistent identifier is the Digital Object Identifier (DOI) (Jacobsen et al., 2020).

F2: *Data are described with rich metadata* (GO FAIR, n.d.-b) (further described in **R1**). The principle is concerned with the discoverability of a resource through searching and filtering, which is enabled by providing descriptions of the resources with rich metadata. A challenge with F2 is the lack of defined minimal "richness" of the metadata, which makes the task of communities to supply sufficient descriptors to their own metadata a considerable difficulty (Jacobsen et al., 2020).

F3: *Metadata clearly and explicitly include the identifier of the data they describe* (GO FAIR, n.d.-b). Principle F3 states that any digital resource description needs to contain the identifier of the resource in question. This way, if the resources and metadata are stored independently, they remain linked. An identified challenge with this point is the selection of machine-actionable metadata model which links the metadata and the resource (Jacobsen et al., 2020).

F4: *Metadata are registered or indexed in a searchable resource* (GO FAIR, n.d.-b). The searchable resource supplies the infrastructure needed for the discovery of metadata record by either utilising the attributes of the metadata (F2) or the identifier of the resources (F3). Numerous challenges are associated with this, such as the lack of a single-source for search which supports all possible metadata fields in all domains, the lack of a uniform way to perform a search which requires multiple software for each tool search, and the limitation of search engines which forbid automated searches, making their application in FAIR-enabled software not possible (Jacobsen et al., 2020).

5.2.1.2.2 Accessible

The second step of FAIRification is about making data Accessible by the users once it is found, and it may include authentication and authorisation. The following sub-steps are utilised:

A1: *(Meta)data are retrievable by their identifier using a standardised communications protocol* (GO FAIR, n.d.-b). A1 describes the ability to retrieve the (meta)data record using a clearly defined mechanism. To enable a fully automated access, the identifier (F1) should adhere to a globally accepted layout, tied to a standardised communication protocol, which supplies a predictable resource access to the user (Jacobsen et al., 2020).

> **A1.1:** *The protocol is open, free, and universally implementable* (GO FAIR, n.d.-b). The sub-principle is concerned with the access process and implies that the protocol and mechanism employed for accessing the resources should not pose an obstacle to the users. The challenge with this is to explicitly document protocols that are not open or free (i.e., such that grant access after personal contact is made) and make them available for the machine-readable metadata. The current solution of this issue is to utilise standardised communication protocols that are free, open, and universally implantable, such as the HTTP protocol (Jacobsen et al., 2020).
>
> **A1.2:** *The protocol allows for an authentication and authorisation procedure, where needed* (GO FAIR, n.d.-b). The sub-principle implies that FAIR does not mean "open", and that additional access measures are required in the instances where restricted digital resources are involved. The current choice is for communities to use protocols when controlling the access of agents to the resources, which protocols should be both as domain specific as needed and as generic as possible (Jacobsen et al., 2020).

A2: *Metadata are accessible, even when the data are no longer available* (GO FAIR, n.d.-b). In the instances when data is no longer available for one or another reason, it is important for agents to have access to the well-described metadata linked with the unavailable data. This is because the data may have been already used and referenced, and access to the metadata ensures the understanding of the historical metadata record and the nature of the inaccessible data. The current challenges with the implementation of this principle are the definition of persistent metadata policies that enable sufficient data description even without its presence, the selection of machine-actionable templates for the persistent metadata policy documentation, and the definition of a machine-actionable scheme to reference it (Jacobsen et al., 2020).

5.2.1.2.3 Interoperable

Data often needs to be integrated with other data; thus, the data needs to be Interoperable with applications, tools, and workflows for storage, analysis, and processing. To ensure that, the following three steps are applied:

I1: *(Meta)data use a formal, accessible, shared, and universally applicable language for knowledge representation* (GO FAIR, n.d.-b). The principle aims to achieve a "common understanding" of the digital resources by using a globally used machine language. Communities must either select an appropriate available technology or come up with a new solution in line with the principle. The selected method should ensure that each data item is the same in multiple resources and interpreted in the exact same way by any agent, be that human or computer agents. The relation of items across the resources needs to be also universally understood. Currently, the most widely used solution which is in-line with the I1 principle is the application of the Resource Description Framework (RDF), which is applied to describe knowledge on the Web in a machine-accessible format (Jacobsen et al., 2020).

I2: *(Meta)data use vocabularies that follow the FAIR principles* (GO FAIR, n.d.-b). This principle is directly linked with principle I1: it requires that the vocabulary used in the knowledge representation language provides sufficient machine distinction of the terms to eliminate false agreements and disagreements. For instance, simple "labels" may be insufficient to enable a machine to understand the contents of the label and the contexts with which it can be properly linked – the simple label "temperature" does not provide sufficient understanding whether the matter in question is "body temperature", "melting temperature", or "outside temperature". Currently, communities need to ensure they utilise terminology systems, units of measure, and classifications which adhere to the FAIR Principles (Jacobsen et al., 2020).

I3: *(Meta)data include qualified references to other (meta)data* (GO FAIR, n.d.-b). An important aspect of FAIR is that the data and metadata do not exist in a silo and are appropriately interlinked with other resources connected to them, which enables the formation of an interlinked network of data and services. The term '*qualified reference*' means a reference to another resource where the relation between the sources is also specified (Jacobsen et al., 2020).

5.2.1.2.4 Reusable

To achieve the main objective of FAIR, which is Reusability, data and metadata should be well-described to enable their replication and combination in different setting, which is achieved by applying the following:

R1: *(Meta)data are richly described with a plurality of accurate and relevant attributes* (GO FAIR, n.d.-b). This principle resembles principle F2 ("*Data are described with rich metadata*"); however, the objective behind F2 is to enable the efficient search and discovery of resources, whereas the rationale of R1 is to enable the assessment (both by humans and machines) of the reusability of the already discovered resource based on their suitability for purpose (Jacobsen et al., 2020).

R1.1: *(Meta)data are released with a clear and accessible data usage license* (GO FAIR, n.d.-b). This principle implies that all data and metadata must always include a license which describes the conditions under which they can be used. It also suggests that if an agent is unable to find the related license, the resources cannot be legally used as well; thus, the lack of licenses does not suggest that the resources are "open" but rather is likely to prevent their reuse. Additionally, data and the related metadata may have different licenses. Currently, there is no well-defined method applied for the distinguishing of a license related to the data and such related to the metadata, leading to broad interpretations of what can and cannot be legally accessed. Thus, communities need to select the most appropriate licenses and licensing requirements for the data as well as for the metadata (Jacobsen et al., 2020).

R1.2: *(Meta)data are associated with detailed provenance* (GO FAIR, n.d.-b). Providing provenance information aids the assessment, both by humans and machines, of whether a discovered resource meets their criteria. Provenance itself includes diverse information about the resource, such as how, why, and by whom the resource was generated, what the conditions were, who the data owner is, if there are any cleansing processes applied post-generation, and more. One of the workable solutions for providing detailed resource provenance is for communities to select a set of metadata descriptors to optimise the provenance, enrich the description, and increase the reusability of the resources (Jacobsen et al., 2020).

R1.3: *(Meta)data meet domain relevant community standards* (GO FAIR, n.d.-b). Multiple disciplinary communities have defined information community standards and adhere to certain best practices. These standards often describe the minimal set of metadata items and description required to assess the quality of the resource. This implies that greater interdisciplinary reusability will require richer metadata. Communities need to decide which data and metadata practices to adhere to, taking into consideration the inter-domain interoperability requirements and deciding which domain-specific requirements should be additionally considered (Jacobsen et al., 2020).

The fifteen FAIR guiding principles mentioned above do not provide technological solutions and are open for interpretation, which may result in inconsistencies and lead to incompatible implementations between stakeholder communities (Jacobsen et al., 2020).

5.2.1.2.5 FAIRification and security

The objectives of FAIR to make (meta)data Findable, Accessible, Interoperable, and Reusable implies that high level security and privacy measures need to be in place, especially in the instances when private and sensitive information is accessed and processed. A study by Delgado and Llorente (Delgado & Llorente, 2020) focuses specifically on the security and privacy aspects of FAIRification in the context of health data. The paper identifies the following methods as steps that can be undertaken to achieve FAIRification in the aspect of security and privacy: pseudonymisation, anonymisation, de-identification, and license attribution. Data anonymisation and de-dentification enable data sharing without compromising the data subjects' privacy. Amongst the approaches that can be utilised for de-identification is dropping data elements from the dataset (Delgado & Llorente, 2020). On the other hand, ISO 25237:2017 on Pseudonymisation (*BS EN ISO 25237:2017: Health Informatics. Pseudonymization*, 2017) defines de-identification as "*the general term for any process of reducing the association between a set of identifying data the data subject*" rather as a specific process itself. ISO 25237:2017 outlines that pseudonymisation enables longitudinal consistency, and allows the association of records, for instance, with each other and keeping them under a pseudonym (*BS EN ISO 25237:2017: Health Informatics. Pseudonymization*, 2017). Pseudonymisation can also be intentionally reversible or irreversible. The same standard also defines anonymisation as the process where, in contrast to pseudonymisation, no longitudinal consistency is provided whilst reducing the possibilities of linking the subjects with their data. Some of the tools utilised by the process include removal, redaction, blanking, randomisation, and substitution of data element from a dataset (*BS EN ISO 25237:2017: Health Informatics. Pseudonymization*, 2017).

5.2.2 Artificial Intelligence for Data Processing

The number of threats the population faces nowadays is concerning, and terrorism is a significant one. Ionescu et al. (Ionescu et al., 2020) research the applications of artificial intelligence (AI) for automatic person and object identification (eProfiler), retrieval of speech intelligence (eTalk), and dissimulated behaviour analysis (eSeeming) in the context of counterterrorism.

5.2.2.1 eProfiler

The eProfiler uses the existing network of video surveillance cameras to provide automated analyses of the presence of persons. The systems integrate three interconnected modules – person/object detection and identification, violence detection, and crowd behaviour analysis.

The first module – person/object detection and identification – allows the detection of a person from a single query image and enables an automated search of their presence on all the video recordings (Ionescu et al., 2020). The initial query image is selected by an operator, and the recognition is bi-level by analysing both facial and body features. The module is capable to detect and recognise potentially threatening objects and track their occurrence on all recording in a comparable manner as with persons. Ionescu et al. (Ionescu et al., 2020) utilise a three-stage detector for this module, including: i) a deep neural network (DNN) which detects the persons/objects; ii) a second network which learns the features at interclass level, and iii) the identification network learns the interclass- characteristics.

The second module of the eProfiler is concerned with violence detection, and it performs real-time analyses of the video from the network of cameras and automatically detects the moment when physical violence occurs by deploying temporal DNNs. The module can predict the occurrence of violence a few minutes before it happens by analysing the behaviour of the persons. The module is based on content descriptors, such as optical flow, and classifiers, such as support vector machines, and deep network architectures (Ionescu et al., 2020).

The third eProfiler module, the crowd behaviour analysis, works similarly to the violence detector by performing real-time analyses of the video and automatically assessing it to detect the formation of a crowd and recognise an abnormal crowd behaviour. The objective of the module is to predict crowd formation and event escalations within the crowd by analysing abnormal behaviours. The module is based on histograms of oriented gradients (HoG) for human detection, support vector machines, hierarchical analysis of groups, and intergroup relations based on optical flow (Ionescu et al., 2020).

5.2.2.2 eTalk

The eTalk system uses the microphone-recorded data along with visual data from cameras and integrates four interconnected modules. The main objective of the system is to retrieve verbal information from potential suspects, and it is designed to automatically process speech in natural language (in this instance – Romanian).

The first module of eTalk automatically translates the recorded speech into text by using algorithms based on Kaldi pipeline with extraction, acoustic, phonetic, and linguistic modelling (Ionescu et al., 2020). The module aims to retrieve, localise specific words, and index the audio data.

Spoken Word Search is the second eTalk module and enables searches in an audio recording. It allows the automatic search of keywords and phrases and when they were used. The module initially utilises the speech-to-text module, and then, direct word search, lemma-based search, and dynamic time warping-based approaches are applied (Ionescu et al., 2020).

The third module of the systems is the speaker identification and validation, which contains two tools: one that identifies the person speaking, and one which performs a validation by checking whether the speech belongs to the identified person, which allows to verify whether audio data was forged. The solution is based on SincNet architectures enhanced with ResNet (Ionescu et al., 2020).

The final module of the eTalk system is the Lip reading, which enables the translations of lip movements into text if the audio information is not available or of insufficient quality. The utilised algorithm consists of two steps, the first one being the localisation of the lips of the speaker by using a neural network, and the second one being the recognition of the lip movements in accordance with pre-defined alphabet of words by using another network (Ionescu et al., 2020).

5.2.2.3 eSeeming

The final system, eSeeming, is applied in a scenario where a discussion between two individuals is happening and has the aim to collect strategic information by integrating two interconnected modules. The system uses information from cameras and microphones which are hidden from the subjects and aims to detect whether the subjects are lying and, thus, validate acquired information.

The first eSeeming module is Emotion analysis, and it provides automatic analyses of the emotional state of the subject by using a recording of their face. The module undertakes three steps: i) estimates the facial action unit's activation; ii) detects the presence of an expression; and iii) detects common emotions, such as sadness and joy. The module is based on semi- and fully supervised DNNs. The same module also performs emotion analyses of the subjects' speech from recordings. The processing steps involved with the voice emotion analysis are, firstly, a pre-processing is performed via normalisation and Hamming windowing; secondly, the features are extracted via Gabor filters; and thirdly, learning is achieved via DNNs (Ionescu et al., 2020).

The final developed module by Ionescu et al. is a tool used for the assessment of some physiological signals assessment of the subjects by performing real-time analyses of their heart rate and respiratory frequency. Dramatic changes in the heart rate and the respiratory frequency may indicate a change in the emotional state of a subject, an instance being the changes that occur when an individual lies. The algorithm only uses visual inputs, and it automatically detects the forehead of the subject from the video records and then performs a plethysmography (changes in volume in different body parts) by frequency analysis which detects the heart rate pulsations and respiratory rhythm (Ionescu et al., 2020).

The researchers report accuracy levels of the developed prototype systems in the range between 70% and 99%.

5.2.3 Data Protection Measures

As of Article 25 of the GDPR (EU General Data Protection Regulation (GDPR), 2016), appropriate security measures should be in place in order to provide data protection "by design and default" (Hansen, 2016). Providing "privacy by design" by definition means that a system is designed in consideration with the privacy and security measures, which effectively become a built-in segment of the system rather than an add-on (Hansen, 2016). Some of the methods that may be applied to ensure data security include encryption, hashing, and tokenisation.

5.2.3.1 Encryption

In broad terms, encryption is the process of encoding information to prevent unauthorised access ("Trusted Data Sharing Framework", 2019). Data encryption encrypts the information exchanged between users or organisations, effectively guaranteeing user safety. The Data Encryption Standard (DES) was developed by IBM, and it is the first recognised cryptographic algorithm (Zhou & Liu, 2022). Generally, there are two primary types of encryptions – the Data Encryption Standard and the Rivest-Shamir-Adleman (RSA).

The DES is a symmetric encryption algorithm, also known as shared key encryption, which uses the same algorithm and keys to encrypt and decrypt data. The process starts with 64-bit plaintext (easy to understand information content) which is divided into left and right 32-bit blocks. After that, 64-bit ciphertext (unrecognisable code derived from plaintext processing) is obtained by applying 16 iterations, cyclic shift transform, and inverse transform (Zhou & Liu, 2022). The corresponding to the ciphertext key is used to decrypt the information. This symmetric encryption benefits from being fast, adaptable to various cryptosystems, and with relatively short keys. However, DES generally does not provide authentication of information integrity, and the key should be kept confidential as its distribution might be dangerous (Zhou & Liu, 2022).

The RSA is an asymmetric encryption algorithm, also known as public key encryption, where a pair of keys is used – one key is used only for encryption, and the other corresponding key is used for decryption (Dhakar et al., 2012). In this algorithm, any authorised user obtains a key pair to communicate with any number of the other parties without exchanging secret keys (Dhakar et al., 2012; Zhou & Liu, 2022). The asymmetric encryption model benefits from having a large key space, but it is slower and less efficient than the symmetric encryption and uses more complex algorithms (Zhou & Liu, 2022).

5.2.3.2 Hashing

Hashing generates a string of fixed length from a text using different algorithms, such as MD5, SHA, and SHA-2. Minor changes on the body of text result in a great difference on the generated hash, making the conversion of the generated hash back to text impossible (“Trusted Data Sharing Framework”, 2019). Hashing algorithms do not require access keys, as compared to encryption, which minimises the risks of unauthorised access in case the key in question is stolen. Hashing is relevant for secure data storage, as the hashed data is unreadable and can be stored safely (“Trusted Data Sharing Framework”, 2019).

5.2.3.3 Tokenisation

Tokenisation is a cryptographic process where sensitive data elements are substituted with non-sensitive elements, or tokens. The tokens do not possess any exploitable value and are only used as identifiers to map the original data. Tokenisation protects both data in transit and stored one (“Trusted Data Sharing Framework”, 2019). The application of this process of data masking prevents the reidentification of the data subjects but can still be identified as uniqueness – the token of a said name, for instance, will always have the same value and shift other places where that name is present to the value (Tachepun & Thammaboosadee, 2020). Table 8 shows a demonstration of Tokenisation.

Table 8: Demonstration of Tokenisation (Tachepun & Thammaboosadee, 2020)

Original	Tokenised
John	Okln
Jane	Olim
Joe	pLuu
John	Okln
Jib	BoFt

5.2.4 Data Storage

5.2.4.1 Forms of Data Storage Discussed

Data storage can take place in various forms, both as a physical entity and as a virtual space. Data can be stored on devices such as hard drives and USB sticks. However, usually data is stored in two separate ways, on-site servers, and cloud-based storage. **On-site storage** provides the users a sense of control over how data is stored and used (Stobierski, 2021). On-site storage can benefit an organisation as data is stored onsite, the stored data can be accessed even without an internet connection. **Cloud storage,** on the other hand, is all online and requires an internet connection. A benefit of cloud storage is that an organisation can continuously purchase more cloud storage space if necessary. The responsibility of cloud-based storage is often with the owner of the system, rather than the customer. The Cloud Service Provider (CSP) is the company responsible for the storage of data, the CSP must "provide access to the data, and the data can't be read or modified by unauthorised users" (Prajapati & Shah, 2022).

A **hybrid cloud architecture** refers to the "composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardised or proprietary technology that enables data and application portability" (Mell & Grance, 2011). P. The combination of the public and private clouds is connected using an encrypted network connection by a virtual private network (VPN) (Manoj Hirway, 2018) and works on the concept that "core activities are hosted on a private cloud, while less essential services are outsourced to a public cloud. Each of the cloud remains a unique entity but linked together by standardised technology" (Odun-Ayo et al., 2018).

A **hyperscale cloud** "refers to the complete mix of hardware and facilities that can scale a distributed computing environment up to thousands of servers. Hyperscale infrastructure is designed for horizontal scalability and leads to high levels of performance, throughput, and redundancy to enable fault tolerance and high availability. Hyperscale computing often relies on massively scalable server architectures and virtual networking" (Hewlett Packard Enterprise, n.d.).

5.2.4.2 UK Defence Storage Solutions

The UK Ministry of Defence has developed a hyperscale cloud service **MODCloud** (see also 4.2.1.5), which is "a public cloud platform for defence which hosts applications, data and related services" (Ministry of Defence, n.d.) and allows users to "control operating systems, storage and deployed applications" (Ministry of Defence, n.d.). Currently the MOD is revising the MODCloud '*technology core*' as it is "fragmented, fragile, insecure and obsolescent" (Ministry of Defence, 2023). The MOD also hopes to further expand MODCloud to build a managed private cloud within the 'SECRET' classification which can be applied across the entirety of His Majesty's Government (Ministry of Defence, 2023).

The UK's Ministry of Defence (MOD) released an outline for the MOD's vision for a **Data Strategy for Defence** in 2021. In this plan, the MOD describes the approach that will be undertaken "the delivery of a multi-classification, multi-cloud and hyperscale cloud environment". By 2025 the MOD aims to use hyperscale cloud services as a platform to exploit emerging technologies, as well as for a foundation on which to build capabilities in big data, advanced analytics, automation, and synthetics (Ministry of Defence, 2023). An aspect of the Data Strategy for Defence is the development of a *Digital Backbone* which should encompass a "singular, secure, modern and digital environment" (Ministry of Defence, 2021) which delivers "a multi-classification, multi-cloud and hyperscale Cloud environment" (Ministry of Defence, 2021). The MOD has set itself a target to be "cloud native as much as possible" which would result in most defence data being stored on the cloud, rather than stored physically.

5.3 Data Security, Sizes, And Storage in The Experts' Organisations

The experts were asked to share the requirements for minimal and maximal data size, methods of data storage, and methods to maintain security in their organisations.

For data size requirements, *43.3%* of the experts shared that there are no specific requirements implemented in their organisations regarding the minimal and maximal data size. For *26.7%* of the organisation's experts belonged to, the size requirements vary based on the data type. *13.3%* said that it is not applicable in their instance. Data size between 1mb and 5mb is required by *3.3%* of the organisations, and the same percentage (*3.3%*) reported a maximum data size requirement of 700kb. The following answers were detected in *3.3%* of the answers: the experts are not aware of such requirements, the experts did not answer the question, and the experts said there is a minimum and maximum data size requirement without specifying the size.

Table 9 shows the data types listed by the experts in regards data storage used in their organisations and projects. The highest-ranking type of data storage in the experts' organisations is on own server (*18.4%*), followed by storage of anonymised data on encrypted folders (*15.8%*). The third most used type of storage in on a Cloud (*13.2%*), followed by unspecified on-site storage (*10.5%*), and the fifth is data storage on a central repository (*7.9%*).

Table 9: Types of data storage implemented by the experts' organisations

Type of data storage	Implemented by % of experts' organisations
Own server	18.4%
Encrypted folders & anonymisation	15.8%
Cloud	13.2%
On-site	10.5%
Central repository	7.9%
N/A	5.3%
Data warehouses	5.3%
Different data spaces for different data types	2.6%
Relational databases	2.6%
Police databases	2.6%
On-site & cloud	2.6%
Internal repository	2.6%
Storage devices	2.6%
Distributed data networks	2.6%
Specific communication system (e.g., Europol's SEINA)	2.6%

Oracle database	2.6%

Table 10 shows the security measures used in the organisations of consulted experts. The top-five ranked methods are: Access control (*21.5%*), Data encryption (*16.9%*), Data anonymisation (*13.8%*), and Adherence to international standards (*4.6%*), as *7.7%* of the experts did not answer this question.

Table 10: Data security measures implemented by experts' organisations

Data security maintenance method	Implemented by % of experts' organisations
Access control	21.5%
Encryption	16.9%
Anonymisation	13.8%
Not answered	7.7%
Adhere to international standards	4.6%
IT solutions	4.6%
Physical security	4.6%
Data backup and recovery	4.6%
Network firewalls	3.1%
Policies & services for authentication	3.1%
Risk management	1.5%
Impact assessment	1.5%
E-ITS (national info security standard)	1.5%
Centrally coordinate data access	1.5%
Credentials-based access	1.5%
System locks based on data security certificates	1.5%
Standard network security access	1.5%
Data retention policies	1.5%
Cut-off internet network	1.5%
Pseudonymisation	1.5%

A variety of mechanisms were described by the experts when asked what the core mechanisms used for data access, transfer, storage, and co-creation in their projects/organisations are.

Table 11: Core mechanisms used for data access, transfer, storage, and co-creation by the consulted experts

Access	Transfer	Storage	Co-creation
• Secured API	• Encryption	• Encryption	• Wikis
• MFA	• Secure communication channels (E.g. VPNs)	• Databases	• Online document editors
• Role-based access control mechanisms	• FTP	• File system	• Version control systems
• Security protocols	• SFTP	• Cloud storage	• Project management software
• Authentication	• HTTP	• Microsoft security measures	• Document sharing platforms
• Authorisation	• HTTPS	• On-premises storage	• Messaging apps
• Encryption	• Data integration platform	• Centralised data repositories	• Microsoft security measures
• Microsoft security measures	• Secure data sharing platforms		• PETs, especially MPC
• Cloud	• Microsoft security measures		
• Logging	• Cloud		
	• Communication systems		
	• Messages transfer frameworks (e.g. JMS)		
	• Nginx server		
	• Web services		
	• APSIs		
	• Various channels		
	• X-Road		

Additionally, *6.7%* of the partners did not answer this question, and *10%* marked it was not applicable in their instance. *20%* of the consulted experts stressed over the importance of accurate data retention and storage, and how improperly maintained data leads to “poisoned-well data” which nobody wants to use. *13.3%* of the experts did not specify the utilised mechanisms but noted that *Only secure data is transferred/Data is accessed only by authorised people/They adhere to the GDPR provisions/Utilise standard IT tools.*

6 Enablers, Barriers and Risks for LAGO RDE Efficiency and Adoption

Section 6 provides a review of the enablers and barriers that, according to the group of consulted experts and documentation, either support or hinder the adoption and efficacy of an RDE for the FCT domain. This section further collates the specific risks identified by experts that might endanger the long-term viability of an FCT RDE. These aspects need to be addressed – comprehensively and in concert – within the LAGO architecture to ensure the viability and efficient functioning of any instantiation of a LAGO RDE.

The nature of enablers and barriers which emerged from the consultations cover a broad range of issues. For easier reference, they have been clustered according to their core focus (infrastructure, data, governance, organisation, etc.). As becomes apparent in the overview, enablers and barriers are often mirror images of each other, meaning that removal of barriers will enable organisations to be more effective in their data practices and collaboration. We nonetheless list the aspects in both enablers and barriers to preserve the specifics of experts' answers and thus a comprehensive view on the information received and reviewed.

6.1 Technological and Data Considerations

Main technological and data enablers:

- Tested and easy to use tools are available on easy to access platforms
- Availability of data exchange software
- Availability of Data privacy/ Privacy Enhancing Technologies (PETs) to safeguard data sharing
- Access to secure technical infrastructure and tools (e.g., data platforms, advanced analytics software, data visualisation tools)
- Efficiency improvements expected due to usage of AI algorithms
- Existence of interoperable data infrastructures/platforms which can facilitate the seamless integration and exchange of data across different systems
- Existence of standardisation of data formats and metadata.

Main technological and data barriers:

- Costs for tools and datasets
- Complexity of technologies
- Lack of anonymisation tools
- A lack of technological infrastructure for sharing/processing data safely
- Regulations that require that "tools need to be publicly auditable" (UK DARE)
- With increased data sharing, there is a risk of data breaches or cyber-attacks, which would result in data loss, data manipulation, or data theft.

6.2 Organisational Considerations

Main enablers within organisations:

- Implementation of training programs on data sharing best practices.
- Education and demonstration of the benefits of data sharing in the FCT domain.
- Clearly defined standards, regulations, and processes within organisations.
- Establishing common standards and protocols to harmonise data handling/processes.

Main barriers within organisations:

- Lack of skilled personnel: CT research requires personnel with specific skills, including expertise in data science, machine learning, statistical analysis, cybersecurity, and domain-specific knowledge in FCT.
- Lack of clear standards and regulations with organisations to guide data sharing and/or collaborations.

6.3 Economic Considerations

Main economics enablers:

- Need for sustainable funding to ensure adequate resources in the development and implementation of the RDE.
- Sufficient and strategic funding for organisations involved in FCT research.

Main economics barrier:

- Inconsistent funding for RDE-related efforts due in part to competing initiatives on a national and EU level.

6.4 Professional and Cultural Considerations

Main cultural enablers:

- The application of appropriate contracts to increase cross-stakeholder trust and reduce cultural barriers.

Main cultural barrier:

- Different stakeholders have varying professional standards and backgrounds.
- Communication gap between security practitioners and technology providers.

6.5 Policy and Governance Considerations

Main policy/governance-related enablers:

- Need for effective governance and data management frameworks.

- A clear policy framework is necessary to ensure that FCT research is conducted in compliance with applicable laws and regulations, and to establish guidelines for data sharing and use.
- Having well-defined legal and policy frameworks is crucial to ensure the secure and ethical sharing of FCT data. This includes policies around data privacy and protection, ethical standards, and guidelines for data anonymisation, sharing, access, and management that are made readily available, easy to understand, and widely disseminated.
- Implementation of well-defined policy rules can increase the trust between researchers and LEAs.
- Strong governance structure and robust data management practices to support secure and effective sharing of FCT data. This includes data sharing agreements, information exchange protocols, and standardization of data formats and definitions.
- Better-informed policy decisions can be achieved when trustworthy and credible research outcomes are produced by ensuring the accuracy, reliability, and impartiality of the data.

Main policy/governance-related barriers:

- Different legislations in different countries (including usage of AI tools/services) hinder effective collaboration.
- The lack of clear governance for managing data sharing/co-creation can lead to confusion and mistrust.
- The lack of applicable policies can lead to disorganisation and mistrust.
- Competing projects on a national/EU level results in duplication of efforts.

6.6 Risks to the Creation and Long-Term Viability of the RDE

The following overview comprises findings from the expert consultations on the key risks that may jeopardise the creation and long-term viability of an RDE in the FCT domain, and which require due consideration. The risks highlight aspects, developments as well as events that could jeopardise the short- and/or long-term success of an FCT RDE. It is vital that the risks are evaluated in advance of formulating core recommendations for the successful and implementation adoption of the RDE.

6.6.1 Architecture Related Risks

- Difficulties in implementing a unified technical solution.
- Technological failures need be taken into consideration. It is a possibility that the technology may not be available to support the development and maintenance of an RDE which would ultimately hinder its success.
- The lack of standardization in data collection, storage, and analysis could lead to inconsistencies and errors in the data, which would undermine the reliability of research findings.
- Different data formats and platforms can hinder the interoperability of data sharing, making it difficult to combine and analyse data from different sources. It was highlighted that central components for data space must be interoperable with national systems and that risks might involve the following: difficulties in data integration, data access and data sharing as a result.
- If the environment is not “easy-to-use" it will result in data sharing and exchange of the data that is unnecessarily complex and opaque.

- Inadequate collaboration in the design and implementation with law enforcement agencies, policymakers, and other stakeholders can result in a misalignment with their needs and priorities, leading to research outcomes that lack practical usefulness.

6.6.2 Data Related Risks

- Potential data leaks resulting from an RDE can jeopardise future projects, as stakeholders would lose faith and trust in these kinds of initiatives.
- Poor data quality can lead to erroneous conclusions and waste resources. It is important to establish protocols for data validation and quality control to ensure the reliability of data.

6.6.3 Resourcing

- The lack of skilled personnel with necessary competencies to support the development and maintenance of the RDE, hindering its success.
- Lack of institutions that make data available leads to low usage/viability of the RDE.

6.6.4 Legal and Ethics Risks

- FCT research may involve sensitive data that is subject to legal and ethical regulations. Failure to comply with these regulations can lead to reputational damage, legal penalties, and loss of funding.
- A possible violation of human rights in general coming from an RDE.
- Legal changes with respect to the data generally, the FCT domain or specific technologies (e.g., AI-related applications).

6.6.5 Trust-Related Risks

- Experts highlighted that a lack of trust between stakeholders often makes it difficult to share data and collaborate effectively. The lack of trust is currently evidenced by the existence of data silos, the duplication of efforts across various domains, and missed opportunities for valuable insights. However, some experts indicated that the observed unwillingness of some organisation to share data may remain despite efforts to create trust, possibly due to organisational and cultural elements.
- Experts highlighted that if LEAs do not understand the benefits in terms of delivering results, then they will not participate fully in the RDE. LEAs appear to be the gatekeepers of FCT data. Therefore, it is vital to demonstrate to LEAs the advantages of having realistic operational data to train AI algorithms, to better fit and meet their needs. LEAs need to be made aware of the benefits of participation in the RDE and unless this is adequately communicated then there will be significant issues in ensuring collaboration.
- A lack of transparency in data sharing, data analysis, and research methodology could erode the credibility and dependability of the research outcomes.

7 Gaps in Current Knowledge

7.1 Best Practices

The FCT research domain requires its own specific set of practices and guidelines. This is lacking at present, and it is not clear what guidelines/practices are commonly employed by all stakeholders when conducting FCT research. Future research will be required to address this gap and should focus on developing best practices and guidelines for data management in the FCT domain. It should also consider the needs and requirements of FCT practitioners when developing tools and systems to support these practices and all types of stakeholders in this field.

7.2 Infrastructure and Resource Requirements

- **Publicly available Resources:** There are insufficient publicly available resources on FCT data sharing. When conducting this research, it was evident that information on FCT data practices is not easily identifiable.
- **FCT Specific resources:** There is a lack of FCT specific policies and data governance frameworks to ensure the creation of FCT specific common standards, regulations, and vocabularies. Presently, experts refer to several standards and policies which have been developed to regulate data practices. These include the following:
  - **General Data Protection Regulation (GDPR):** established by the European Union to set standards for data protection and privacy, including requirements for data controllers and processors, data subject rights, and data breach notifications.
  - **FAIR (Findable, Accessible, Interoperable, and Reusable) principles:** to ensure research data is easy to discover, access, share, and reuse by providing clear guidelines for metadata and data management.
  - **Open Science Policy Platform (OSPP) guidance:** on best practices for open science, including open access to research data, publications, and software, and transparent and collaborative research practices.
  - **ISO 27001:** This international standard provides a framework for information security management systems (ISMS), which can be used to ensure the confidentiality, integrity, and availability of sensitive data.
  - **Research Data Alliance (RDA) Guidelines:** which provides guidelines and recommendations for data management and sharing, including best practices for data citation, data repositories, and data interoperability.

The nature and sensitivity of FCT research would indicate that this area requires specific focus and would benefit from the creation of a comprehensive guidance package for all practitioners/stakeholders in this field.

7.3 Technological Gaps

After analysing the existing literature and expert consultations, gaps highlighted were:

- **User friendly innovative technologies:** Experts indicated that new tools and technologies are needed to enable the efficient analysis and visualization of large and complex datasets generated in FCT research. Experts highlighted that these tools should be user-friendly and accessible to non-expert users, while also providing advanced capabilities for data exploration and modelling. Experts highlighted that due to low technical competencies it is important that enabling tools are developed and made available.
- **Automation:** Experts indicated that there is currently a need to develop automated tools to assist in analysis and interpretation of FCT data.
- **AI development:** Experts stated that AI development for analysis and visualizations would be of considerable importance. The current status of this area is unknown.
- **Tool Repository:** Experts highlighted the need for a "***trusted***" EU FCT repository of tools (analysis tools, ML-DL tools) for use and testing by LEAs.
- **Storage solutions:** Common European cloud storage infrastructure that is safe and easy to use and can handle big data sizes.
- **Secure Platforms:** Experts indicated that there is a need to develop secure and trusted data sharing platforms. There was little or no knowledge provided on the most appropriate or fit for purpose platforms that might currently be in existence.

7.4 Gaps in Common Policy and Governance of FCT domain

After a meticulous analysis of the current literature and consultations with experts, several gaps in policy and governance were identified as follows:

- **Governance of ethical, social, and legal issues in FCT research domain:** There is no concrete and cohesive guide on how ethical, legal, and social issues relating to the FCT research domain should be governed and managed. *FCT research raises several ethical, legal, and social issues, such as issues of privacy, data protection, and ownership. Future research should address these issues and develop new frameworks for ensuring ethical and responsible conduct of FCT research, as well as for engaging with stakeholders and the wider public.*
- **Cross – stakeholder data sharing practices:** There is no information made available in the relevant documentation shared by the experts about policies concerning cross-stakeholder data sharing and collaboration.
- **National Level Policies:** A gap in the recommended documentation and expert consultations was noted in regard to the national level policies regarding FCT research governance.
- **Policy development in FCT research domain:** There is no information available on policy development in the FCT research domain.
- **Common policies:** There is a lack of concrete information on common policies currently used with regard to data handling in FCT research. Only one relevant policy (The FCT Open Science Policy) is mentioned by one of the experts, suggesting a gap in relevant policies.

- **Trust based governance framework:** A considerable gap is that presently there is no trust-based governance framework in place to encourage collaboration across stakeholder groups in the FCT research domain. The experts indicated that legal know how on how agreements are made at present between LEAs and other FCT domain stakeholders is unclear. The findings from this research also indicates that stakeholders are unclear as to the most appropriate methods to increase trustworthiness and confidence between the LEAs and researcher communities.

7.5 Cultural and Organisational Gaps

- **Transparency:** Experts indicated that at present there is reluctance to consider partaking in collaborations with stakeholders outside of their immediate organisation. This is due to a several concerns mainly relating to trust, training, and security of data.
- **Collaboration and trust building:** Experts indicated that the lack of training on data handling practices could be attributed to limited funding provision. Relevant documentation and expert consultations indicated that trust, be it related to the general population or between organisations, is paramount. However, neither have specified or recommended methods to increase trust other than the creation and implementation of training programs and standardised templates for agreements. The expert specifications provided for what these training programs should entail are vague, but suggestions include training on data sharing best practices and data management skills to raise awareness on the benefits of data sharing within the FCT domain.

7.6 Summary

Table 12 shows the different elements of LAGO FCT RDE (Figure 6) and whether they were addressed by the experts in their consultations as good practices, recommendations for the development of LAGO RDE, or methods utilised in their organisations/projects. The unmarked elements suggest there are gaps in knowledge in the practices currently utilised by experts.

Table 12: Table showing the various LAGO FCT RDE Concept elements and if they were addressed by the experts

Governance		Usage		Design principles	
LAGO FCT RDE Concept	**Addressed by the experts**	**LAGO FCT RDE Concept**	**Addressed by the experts**	**LAGO FCT RDE Concept**	**Addressed by the experts**
Identity	☐	Quality assessment	☒	Decentralisation	☒
Usage control	☐	Compliance assessment	☒	Security and trust	☒
Meta-data	☒	Access	☒	Proportionality and Risk	☒
Vocabularies	☐	Risk assessment	☒	Data sovereignty	☒
Catalogue	☐	Model sharing	☒	Data quality	☒
Versioning	☒	Trusted and secure data testing	☒	Openness	☒

Agreements	☒	Logging / tracing	☒	Interoperability	☒
				Transparency	☒
				Portability	☒
				Ethics, Legal and Privacy	☒

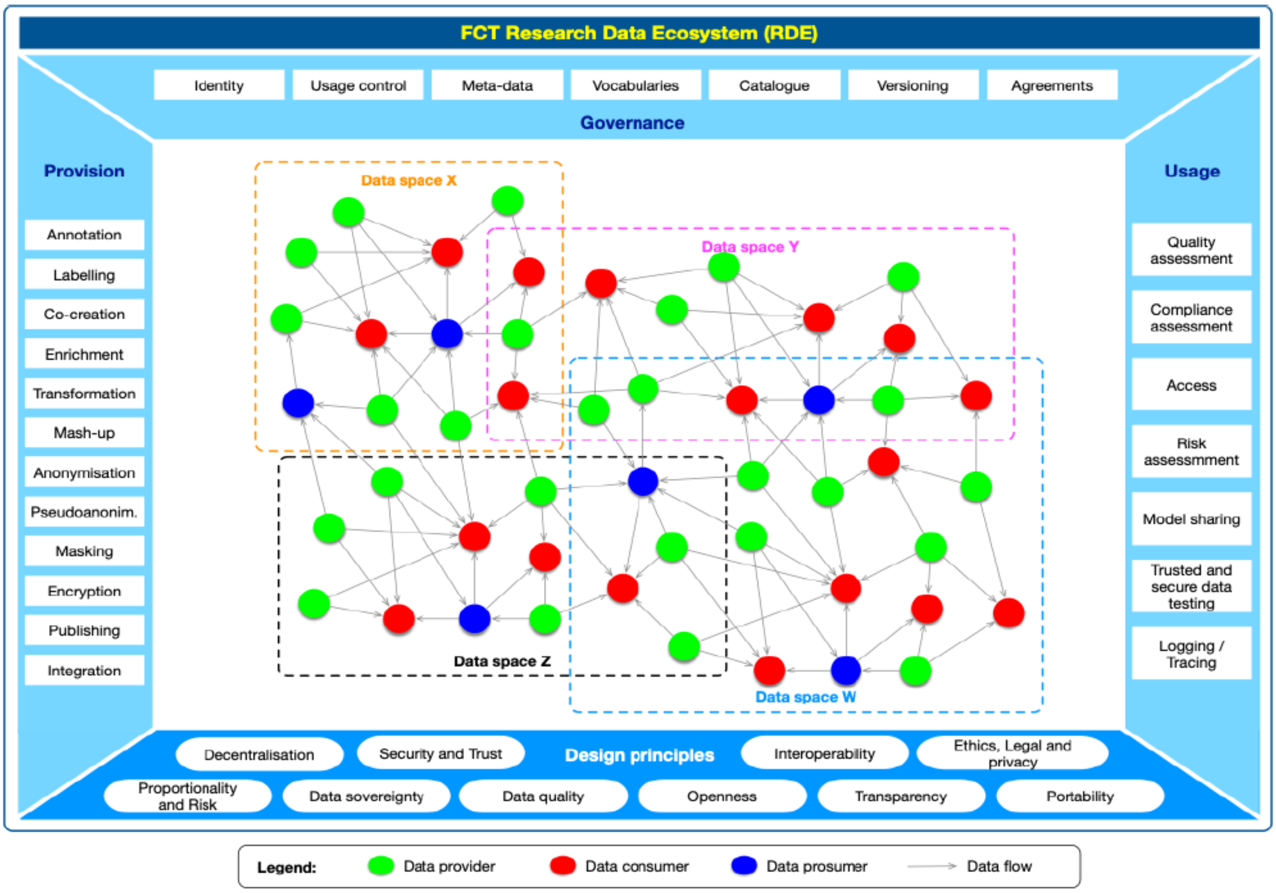


Figure 6: LAGO FCT RDE Concept (derived from LAGO DoA)

8 Recommendations for LAGO RDE

8.1 Effective Data Handling: Experts' Opinion

The experts were asked what they do as individuals to ensure effective data handling in FCT research. This question was not answered by *21.43%* of the experts. *16.6%* of the experts stated they ensure effective data handling by adhering to relevant data security guidelines/grant regulations/organisational rules/data handling policies/data protection laws. Some of the mentioned regulations in this section include ISO27001 and NATO SECRET compliance. Data encryption and secure data storage was answered by *14.29%* of the experts. Access control and training (either the participation in such or their planning for users) were answered by *7.14%* of the experts. The application of proven IT tools only and pseudonymising/anonymising the data are each implemented by *4.76%* of the experts. Each of the following answers was provided by *2.38%* of the consulted experts: *providing organisational support through a common ICT strategy and budget; participation in organisational discussions on data handling; not sharing data without an agreement; application of synthetic data; data verification; development of a platform for data collection, visualisation, and analysis; testing with test datasets; data risk study;* and *physical protection of local data servers.*

Table 13: Effective data handling practices in FCT research implemented by the experts

Effective data handling practice	Implemented by % of the experts
Not answered	21.43%
Implement data handling policies / adhere to data security guidelines / grant regulations / organisational rules	16.67%
Store data securely / encryption	14.29%
Access control	7.14%
Provide / participate in trainings	7.14%
Use of proven IT tools	4.76%
Pseudonymisation / anonymisation	4.76%
Not directly involved	2.38%
Support though a common ICT strategy and budget	2.38%
Organisational discussions related to data handling	2.38%
Not share data without an agreement	2.38%
Creation / use of synthetic data	2.38%
Data verification	2.38%
Platform development for data collection, visualisation, and analysis	2.38%
Physical protection of local data servers	2.38%
Testing with test datasets	2.38%
Data risk study	2.38%

8.2 Recommended Data Standards and Policies

Given experts inputs, a number of standards and policies were recommended viable for data practice regulations in the context of the LAGO RDE. Often experts referred to them in generic terms as standards on inputting data, cyber standards, medical standards or national legislation which should be considered. Table 14 presents the recommendations, together with notes on specific aspects to consider (as and if provided by experts).

Table 14: Standards and policies on data practice regulation recommended by the consulted experts

	Name of recommended documentation	Notes from experts
1	GDPR	Sets high standards for data privacy, security, and transparency for individuals in the EU and EEA.
2	Data Act	Interoperability.
3	ISO/IEC 27001	International standard for information security management that outlines requirements for risk management, security controls, and continuous improvement.
4	FAIR	"Findable, Accessible, Interoperable, Reusable" Principles
5	Research Data Alliance	The RDA provides guidelines and recommendations for data management and sharing, including best practices for data citation, data repositories, and data interoperability.
6	Digital Commons network to promote development of freeware	
7	Data stewardship task force by UNECE	
8	ISO/IEC 19794-5	For face capture
9	ISO/IEC 19794-4	For finger capture
10	ISO/IEC JTC 1 SC 37	Biometrics
11	ISO/IEC JTC SC27	Security Techniques
12	ISO/IEC JTC 1 SC 17	Cards and Personal Identification
13	DCAT	Data Catalogue Vocabulary
14	Green Book on the Machine-Readable Publication of Estonian Public Information	National level, specifically focuses on open data.
15	Universal Message Format	Useful for data exchange
16	Open Science Policy Platform	The OSPP provides guidance on best practices for open science, including open access to research data, publications, and software, and transparent and collaborative research practices.
17	Polish Police unified data collection system	The Polish police is trying to introduce a unified data collection system, but these solutions are not yet worth disseminating and only allow for the preparation of crime statistics.

18	TRUST	"Transparency, Responsibility, User focus, Sustainability, Technology" Principles
19	"Data Governance: A conceptual framework, structured review, and research agenda"	A comprehensive overview for data governance is given by Abraham et al.
21	National classified data protection policies	Not known
22	EU classified data protection policies	Not known

8.3 Best Practices in FCT Research: Experts' Opinion

The experts were asked to give examples from their experience of good practices in FCT research without naming people or organisations. The majority of the experts (*46.7%*) did not answer this question, and *16.7%* stated they do not have such experience. *10%* of the experts listed the replacement of real data with synthetic as a good practice. Each of the following answers was given by *3.3%* of the experts: *the SIRIUS platform by EUROPOL EU IRU; enrichment of user requirements by technical partners showing the potentials of the technologies to the end users; partners ensuring accuracy, reliability, impartiality of data by applying rigorous data collection methods, appropriate statistical techniques, and unbiased interpretation; application of open data repositories which make data accessible to research communities; making data accessible only internally within the research team and granting public access only to the report/analysis; application of data analytics and machine learning; anonymisation of real data and transforming it into fake data;* and *easy to use encryption when sharing data*.

Table 15: Particularly good practices in FCT research reported by the experts

Particularly good practices in FCT research	Reported by % of the experts
Not answered	46.7%
No experience	16.7%
Replacing real data with synthetic	10%
SIRIUS platform by Europol EU IRU	3.3%
Enrichment of user requirements by technical partners showing the potentials of the technologies to the end users	3.3%
Ensuring accuracy, reliability, impartiality of data by applying rigorous data collection methods, appropriate statistical techniques, and unbiased data interpretation	3.3%
Open data repositories to make data accessible to the research community	3.3%
Data is accessible only internally within the research team, and the report/analysis is publicly available	3.3%
Application of data analytics and machine learning	3.3%
Anonymisation of real data and transforming it into fake data	3.3%
Easy-use encryption techniques when data is shared	3.3%

8.4 Policy and Governance

- Adequate collaboration with policy makers is paramount for the success of an RDE in the FCT domain.
- Establishment of clear governance structures for monitoring and enforcing compliance with ethical and legal standards within the RDE.
- Establishing a strong governance structure and robust data management practices can support secure and effective sharing of FCT data. This includes data sharing agreements, information exchange protocols, and standardization of data formats and definitions.

8.5 Organisational: Skills and Standards

- **Training and education**: There is need to ensure that skilled personnel with specific skills involving data science, machine learning, statistical analysis, cybersecurity and domain specific knowledge are involved in FCT. In order to ensure that this is facilitated, organisations will need to participate in, or create, training and education programs that promote and teach data sharing best practices, build on data management skills and raise awareness of the benefits of data sharing. These activities will be key in overcoming current barriers to instilling a culture of data sharing amongst trusted partners and stakeholders.

- **FCT specific manual for practitioners**: FCT research is unique in that it involves security sensitive data. Thus, FCT domain specific research requires a resource which includes all applicable information on legal and ethical standards, regulations, data handling measures and protocols included in one resource.

8.6 Finance

- Information on funding the development and implementation of the RDE needs to be made clear and available to participating stakeholders within LAGO. There is a clear concern around the compatibility of national technical infrastructure and resources to meet the needs of LAGO's research data ecosystem.
- The table below shows the most common requirements for successful FCT research, gathered from the expert consultations.

Table 16: Required Resources for successful FCT research.

Required Resources	
Legal	11%
Finance/Funding	9%
Policy	8%
Standards	6%
Clarity of definitions	5%
Trained personnel	5%
Technical	5%
Template/components for data sharing agreements	5%

8.7 Organisations to be involved in FCT RDE

The consulted experts highlighted 18 different categories of organisations that should be included in the FCT-specific RDE. The majority of the experts listed more than one group in their consultations. The top-most needed to be involved group is LEAs, including specialist functions with LEAs and intelligence entities, which was mentioned by *21.6%* or the consulted experts. The second most mentioned group was researchers with *13.5%*, followed by companies developing AI, applications, and tools (*11.7%*). International organisations (such as EUROPOL and UN) and legal experts were each mentioned by *9%* of the consulted experts. Data providers and scientist and Private sector (such as social media providers and industries) were each mentioned by *7.2%* of the experts. The Ministry of Interior was reported as a group that should be involved in FCT-specific RDE by *6.3%* of the experts, and other governmental organisations by *4.5%*. Policy level experts, non-governmental organisations, and financial institutions were highlighted by *1.8%* of the experts. Subject matter experts, Transport sector, Military, Private security specialists, and Civil society organisations were mentioned by *less than 1%* of the consulted experts.

The experts were also consulted whether there should be any groups not involved in the FCT-specific RDE. On this question, *63.3%* of the experts expressed that no organisations or groups should be excluded from the FCT RDE, as some of the experts suggested granting different levels of data access and implementation of verification. Organisations and companies from non-democratic countries should be excluded from the FCT RDE according to *6.7%* of the experts. The same number of experts (*6.7%*) did not answer this question. Each of the following groups should be excluded according to *3.3%* of the consulted experts: citizens, big technological suppliers, political organisations, private companies that are not licensed to handle the required sensitivity levels of data, and groups that do not have legitimate stake in the FCT domain. Exclusion based solely on sound ethical and legal principles was also mentioned by *3.3%* of the consulted experts.

Some discrepancy can be observed in the abovementioned data. Although *11.7%* of the experts specifically mentioned that companies developing AI-tools, applications, and other relevant tools should be included in the RDE, *3.3%* exclusively list big technological companies as such that should be excluded from the ecosystem.

Following the aforementioned questions, the experts were asked whether the parties they had identified as ones that should be part of the ecosystem have similar frameworks, standards, and strategies on sharing, co-creating, and managing data. On this question, *43.3%* of the experts said that the mentioned organisations have different standards, followed by *16.7%* acknowledging the differences of the current standards of the groups and the need for establishment of common ones. *10%* of the experts' state that the recommended groups have similar standards, frameworks, and strategies. Out of the consulted experts, *6.7%* said that they are not aware of the strategies of the listed organisations, and *6.7%* answered “N/A”. Stated by *3.3%* of the experts were each of the following: similar strategies on a national level, similar strategies for governmental institutions but possibly different for the private industry, and different strategies due to the scale, type and sensitivity of data handled by the groups.

8.8 Lessons Learned in FCT Research Practices: Experts' Opinion

The experts were asked to give examples from their experience of bad practices in FCT research without naming people or organisations. The majority of the experts (*36.7%*) did not answer this question, and *13.3%* stated they do not have such experience. *20%* of the experts stated that a particularly bad practice they experience is partnering organisations not sharing outputs. Each of the following answers was given by *3.3%* of the experts: *social media companies denying access to API; neglecting the need to validate the "privacy by design" requirements; delays in data validation; intentional manipulation of data (falsification, selective reporting, data suppressing) to achieve a predetermined outcome; improper data anonymisation; insufficient data security practices; data misuse for political purposes; multiple individuals accessing data using the same login credentials;* and *LEAs sharing either an abundance of data on security activities or none at all*.

Table 17: Particularly bad practices in FCT research reported by the experts

Particularly bad practices in FCT research	Reported by % of the experts
Not answered	36.7%
Not sharing outputs	20%
No such experience	13.3%
Social media companies deny access to API	3.3%
Neglecting the need to validate the "privacy by design" requirements	3.3%
Data validation delays	3.3%
International manipulation of data (falsifying data, selectively reporting, suppressing data) to achieve a predetermined outcome	3.3%
Improper data anonymisation	3.3%
Lack of data security	3.3%
Data misuse for political purposes	3.3%
LEAs sharing either an abundance of data on security activities or none at all	3.3%
Multiple individuals using the same login credentials	3.3%

9 Conclusions

This document provides a review of the current landscape of regulations, practices, barriers, and enablers for data management in the FCT domain. It offers an important integration of documentation available from existing projects and initiatives, documentation, and experts in the FCT domain. Given the extent of the landscape and the breadth of the analysis the perspectives presented have of necessity limitations. Most importantly, due to language restrictions only a limited number of nation-specific documents were able to be reviewed. Thus, there are likely additional national practices, regulations, standards, policies, and experiences that have not been included. Also, we encountered limitations due to access restrictions to domain-specific documentations, specifically from LEAs.

Overall, however, this deliverable offers an important view on, firstly, the diversity of data practices and requirements by stakeholders towards an FCT-based RDE including good practices, secondly, the requirements and expectations by stakeholders for the effective design of the LAGO RDE. This report should thus be considered a springboard that informs and shapes further engagements throughout the LAGO project in developing the architecture as well as its technological support.

10 References

ADR UK. (n.d.-a). *About ADR UK*. Retrieved March 15, 2023, from https://www.adruk.org/about-us/about-adr-uk/

ADR UK. (n.d.-b). *SafePod Network*. Retrieved March 15, 2023, from https://www.adruk.org/data-access/safepod-network/

Amazon Web Services. (n.d.-a). *AWS Storage Gateway*. https://aws.amazon.com/storagegateway/?nc=sn&loc=0

Amazon Web Services. (n.d.-b). *Types of Cloud Computing*. https://aws.amazon.com/types-of-cloud-computing/

Baron, J., & Kotecha, S. (2013). *Amazon Web Services-AWS Storage Options Storage Options in the AWS Cloud*. http://aws.amazon.com/whitepapers/

Bowie, N. G. (2021). Terrorism Research Initiative 40 Terrorism Databases and Data Sets. *Source: Perspectives on Terrorism*, *15*(2), 147–161. https://doi.org/10.2307/27007301

BS EN ISO 25237:2017: Health informatics. Pseudonymization. (2017). British Standards Institute .

Cirlig, C.-C. (2022). *Strengthening Europol's Mandate* .

Council of the European Union. (2005). *Convention between the Kingdom of Belgium, the Federal Republic of Germany, the Kingdom of Spain, the French Republic, the Grand Duchy of Luxembourg, the Kingdom of the Netherlands and the Republic of Austria on the stepping up of cross-border cooperation, particularly in combating terrorism, cross-border crime and illegal migration.*

DARE UK. (n.d.). Retrieved March 3, 2023, from https://dareuk.org.uk/

DARE UK. (2021a). *Data Research Infrastructure Landscape*. https://doi.org/10.5281/zenodo.5584696

DARE UK. (2021b). *Data Research Infrastructure Landscape*. https://doi.org/10.5281/zenodo.5584696

Delgado, J., & Llorente, S. (2020). Security and privacy when applying FAIR principles to genomic information. *Studies in Health Technology and Informatics*, *275*, 37–41. https://doi.org/10.3233/SHTI200690

Desai, T., Ritchie, F., & Welpton, R. (2016). Five Safes: designing data access for research. *Economics Working Paper Series: University of the West of England*.

Dhakar, R. S., Gupta, A. K., & Sharma, P. (2012). Modified RSA Encryption Algorithm (MREA). *Proceedings - 2012 2nd International Conference on Advanced Computing and Communication Technologies, ACCT 2012*, 426–429. https://doi.org/10.1109/ACCT.2012.74

Domingo-Ferrer, J., Farràs, O., Ribes-González, J., & Sánchez, D. (2019). Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges. In *Computer Communications* (Vols. 140–141, pp. 38–60). Elsevier B.V. https://doi.org/10.1016/j.comcom.2019.04.011

Eesti avaandmete teabevärav. (n.d.). Retrieved March 13, 2023, from https://avaandmed.eesti.ee/

Estonian Ministry of Interior. (2022). *ICT Strategy* . Estonian Ministry of Interior .

Estonian Ministry of the Interior, & Estonian Ministry of Justice. (2021). *Joint action plan for the digitalization of criminal proceedings in the administration of the Ministry of Justice and the Ministry of the Interior* . Estonian Ministry of Interior .

European Commission. (2020). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act)*.

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council on automated data exchange for police cooperation ("Prüm II"), amending Council Decisions 2008/615/JHA and 2008/616/JHA and Regulations (EU) 2008/1726, 2019/817 and 2019/818 of the European Parliament and of the Council*.

European Parliament. (2022). *Personal data protection*.

EUROPOL. (2023). *Data Protection & Transparency: Balancing Europol's operational needs and the individual's right to data protection*.

Genomics England . (n.d.). Retrieved March 14, 2023, from https://www.genomicsengland.co.uk/

GO FAIR. (n.d.-a). *FAIR Principles*. GO FAIR. Retrieved March 7, 2023, from https://www.go-fair.org/fair-principles/

GO FAIR. (n.d.-b). *FAIR Principles*. GO FAIR. Retrieved March 7, 2023, from https://www.go-fair.org/fair-principles/

GOV.UK. (n.d.). *Data Protection*. https://www.gov.uk/data-protection

gov.uk. (2020, February 12). *Research Code of Practice and Accreditation Criteria: Part 2: Accreditation Criteria: Section B: Accreditation of researchers and peer reviewers*. Digital Economy Act 2017. https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria#section-b-accreditation-of-researchers-and-peer-reviewers

Hansen, M. (2016). Data Protection by Design and by Default à la European General Data Protection Regulation. In *Privacy and Identity Management. Facing up to Next Steps* (pp. 27–38).

Hellenic Data Protection Authority. (2022). Επεξεργασία δεδομένων υγείας. In *Hellenic Data Protection Authority* . Hellenic Data Protection Authority . https://www.dpa.gr/el/enimerwtiko/thematikes_enotites/eidikeskatigories/dedomenaugeias/epexergasia_dedomenwn_ugeias

Hewlett Packard Enterprise. (n.d.). *Hyperscale*. https://www.hpe.com/us/en/what-is/hyperscale.html

Ionescu, B., Ghenescu, M., Rastoceanu, F., Roman, R., & Buric, M. (2020). Artificial Intelligence Fights Crime and Terrorism at a New Level. *IEEE Multimedia*, *27*(2), 55–61. https://doi.org/10.1109/MMUL.2020.2994403

ISO 27001 Information Security Management. (n.d.). Retrieved March 13, 2023, from https://www.iso.org/isoiec-27001-information-security.html

Jacobsen, A., Azevedo, R. de M., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2020). Fair principles: Interpretations and implementation considerations. In *Data Intelligence* (Vol. 2, Issues 1–2, pp. 10–29). MIT Press Journals. https://doi.org/10.1162/dint_r_00024

Karakachanov, K. (2018a). Политика за защита на личните данни в Министерство на Отбраната, структурите на пряко подчинение на Министъра на Отбраната и Българската Армия (Policy of Personal Data Protection of the Ministry of Defence and the Bulgarian Army) . In *Ministry of Defence, Republic of Bulgaria* .

Karakachanov, K. (2018b). Политика за защита на личните данни в Министерство на Отбраната, структурите на пряко подчинение на Министъра на Отбраната и Българската Армия (Policy of Personal Data Protection of the Ministry of Defence and the Bulgarian Army) . In *Ministry of Defence, Republic of Bulgaria* .

Kavianpour, S., Sutherland, J., Mansouri-Benssassi, E., Coull, N., & Jefferson, E. (2021a). A Review of Trusted Research Environments to Support Next Generation Capabilities based on Interview Analysis. *Journal of Medical Internet Research* .

Kavianpour, S., Sutherland, J., Mansouri-Benssassi, E., Coull, N., & Jefferson, E. (2021b). A Review of Trusted Research Environments to Support Next Generation Capabilities based on Interview Analysis. *Journal of Medical Internet Research* .

Machado, H., & Granja, R. (2018). Ethics in Transnational Forensic DNA Data Exchange in the EU: Constructing Boundaries and Managing Controversies. *Science as Culture*, *27*(2), 242–264. https://doi.org/10.1080/09505431.2018.1425385

Manoj Hirway. (2018). *Hybrid Cloud for Developers*. https://learning.oreilly.com/library/view/hybrid-cloud-for/9781788830874/

Mell, P. M., & Grance, T. (2011). *The NIST definition of cloud computing*. https://doi.org/10.6028/NIST.SP.800-145

Milieu Consulting. (2020). *Study on the practice of direct exchanges of personal data between Europol and private parties Final Report*. https://home-affairs.ec.europa.eu/system/files/2020-11/publication_study_private_parties.pdf

Ministry of Defence. (n.d.). *Host applications on MODCloud*. https://mod-cloud.service.mod.gov.uk/

Ministry of Defence. (2020a). Defence Data Management Strategy . In *HM Government* . HM Government .

Ministry of Defence. (2020b). Defence Data Management Strategy . In *HM Government* . HM Government .

Ministry of Defence. (2021). *Data Strategy for Defence*.

Ministry of Defence. (2023). *Cloud Strategic Roadmap for Defence*. https://www.gov.uk/government/publications/cloud-strategic-roadmap-for-defence/cloud-strategic-roadmap-for-defence#the-future-of-cloud-for-defence

Monahan, T. (2009). The murky world of 'Fusion Centres.' In *Criminal Justice Matters* (Vol. 75, Issue 1, pp. 20–21). https://doi.org/10.1080/09627250802699715

Odun-Ayo, I., Ananya, M., Agono, F., & Goddy-Worlu, R. (2018, August 17). Cloud Computing Architecture: A Critical Analysis. *Proceedings of the 2018 18th International Conference on Computational Science and Its Applications, ICCSA 2018*. https://doi.org/10.1109/ICCSA.2018.8439638

Open Government Partnership, & Hellenic Republic Ministry of Administrative Reconstruction. (2022). *4th National Action Plan on Open Government 2019-2021 - Updated Version-Including Addendum with Additional Commitments incorporated through the OGP Greece Ideathon Editorial group*.

OpenSAFELY. (n.d.). Retrieved March 14, 2023, from https://www.opensafely.org/

Pickering, J. C., & Fox, A. M. (2022). Enabling Collaboration and Communication Across Law Enforcement Jurisdictions: Data Sharing in a Multiagency Partnership. *Criminal Justice Policy Review*, *33*(7), 732–755. https://doi.org/10.1177/08874034211066756

Prajapati, P., & Shah, P. (2022). A Review on Secure Data Deduplication: Cloud Storage Security Issue. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 34, Issue 7, pp. 3996–4007). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2020.10.021

Protection Information Management. (2018). *Framework for Data Sharing in Practice* .

Puusalu, J. (2020). *SUURANDMED: OLEMUS JA KASUTAMISE KITSASKOHAD*. https://digiriiul.sisekaitse.ee/handle/123456789/2537

SafePod. (n.d.). *SafePod Locations*. Retrieved March 15, 2023, from https://safepodnetwork.ac.uk/safepod/

SAIL Databank. (n.d.). Retrieved March 14, 2023, from https://saildatabank.com/

Stobierski, T. (2021, March 2). *5 Key Elements of a Data Ecosystem*. https://online.hbs.edu/blog/post/data-ecosystem

Stokes, P. (2017a, January 27). *The 'Five Safes' – Data Privacy at ONS*. Office for National Statistics . https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/

Stokes, P. (2017b, January 27). *The 'Five Safes' – Data Privacy at ONS*. Office for National Statistics . https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/

Tacconelli, E., Gorska, A., Carrara, E., Davis, R. J., Bonten, M., Friedrich, A., Glasner, C., Goossiens, H., Hasenauer, J., Abad, J. M. H., Penalvo, J. L., Sanchez-Niubo, A., Sialm, A., Scipione, G., Soriano, G., Yazdanpanah, Y., Vorstenbosch, E., & Jaenisch, T. (2022). Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response. *The Lancet Regional Health - Europe*, *21*.

Tachepun, C., & Thammaboosadee, S. (2020, July 1). A Data Masking Guideline for Optimizing Insights and Privacy under GDPR Compliance. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3406601.3406627

Tech Against Terrorism. (2021). *Transparency Report: Terrorist Content Analytics Platform*.

The European Parliament, T. E. C. (2016, April 27). *EU General Data Protection Regulation (GDPR) Art. 5 Principles relating to processing of personal data* . https://gdpr-info.eu/art-5-gdpr/

EU General Data Protection Regulation (GDPR) , (2016).

The European Parliament, & The European Council. (2016, April 27). *EU General Data Protection Regulation (GDPR): Art. 4 GDPR Definitions*. Official Journal of the European Union . https://gdpr-info.eu/art-4-gdpr/

The Scottish Government. (2015a). *A Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics*. http://www.esrc.ac.uk/files/publications/themed-publications/improving-access-for-research-and-

The Scottish Government. (2015b). *A Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics*. http://www.esrc.ac.uk/files/publications/themed-publications/improving-access-for-research-and-

The Scottish Government. (2015c). *A Charter for Safe Havens in Scotland: Handling unconsented data from National Health Service patient records to support research and statistics* .

Trusted Data Sharing Framework . (2019). In *Infocomm Media Development Authority of Singapore (IMDA) & Personal Data Protection Commission (PDPC)*. https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

UK Department of Health and Social Care. (2022). Data saves lives: reshaping health and social care with data. In *UK Department of Health and Social Care* . UK Department of Health and Social Care . https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

UK Health Data Research Alliance. (2020a). *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data science needs*.

UK Health Data Research Alliance. (2020b). *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data science needs*.

UK Health Data Research Alliance. (2021). *Building Trusted Research Environments: Principles and Best Practices; towards TRE ecosystems* .

UK Ministry of Defence. (2021a). Data Strategy for Defence: Delivering the Defence Data Framework and exploiting the power of data. In *HM Government* . HM Government .

UK Ministry of Defence. (2021b). Data Strategy for Defence: Delivering the Defence Data Framework and exploiting the power of data. In *HM Government* . HM Government .

UK Ministry of Defence. (2021c). Digital Strategy for Defence: Delivering the Digital Backbone and unleashing the power of Defence's data. In *HM Government* . HM Government .

UK Ministry of Defence. (2021d). Digital Strategy for Defence: Delivering the Digital Backbone and unleashing the power of Defence's data. In *HM Government* . HM Government .

UK Research and Innovation, ADR UK, & Health Data Research UK. (2021). *Data Research Infrastructure Landscape: A review of the UK data research infrastructure* .

U.S. Department of Homeland Security. (2022). *Fusion Centers*. https://www.dhs.gov/fusion-centers

U.S. Department of Justice. (n.d.). *Privacy and Civil Liberties, Justice Information Sharing*. https://bja.ojp.gov/program/it/privacy-civil-liberties

Voronova, S. (2021). *Understanding EU counter-terrorism policy* . https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/659446/EPRS_BRI(2021)659446_EN.pdf

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.18

X-Road. (n.d.). Retrieved March 10, 2023, from https://x-road.global/

Zhou, L., & Liu, C. (2022). The improvement of data encryption technology in computer network security. *Proceedings - 2022 International Conference on Artificial Intelligence in Everything, AIE 2022*, 465–470. https://doi.org/10.1109/AIE57029.2022.00095

11 Appendices

11.1 Appendix A – Methodological Framework

Executive Summary

This document intends to outline T3.1 Consensus Report methodological framework and outlines the activities undertaken to effectively achieve the objectives of T3.1 as well as key information regarding the carrying out of such activities, partners' roles, and key deadlines for their development and implementation.

Objectives of T3.1 Research Activities

Lead WP3: ENG **Participants:** all partners

- **To identify:**
  - Best and effective practices
  - Current and future usage scenarios for the data (ties into T3.2)
  - Requirements and factors that impact ability or willingness for data sharing (including trust)
- **Special attention needed in:**
  - Analysis of **existing data systems** and **sources relevant** for FCT research **data formats and mechanisms** for **data transfer, storage, and security**.
  - Identification of **R&D projects for relevant datasets** and **methods for creating them** (AIDA, GRACE, STARLIGHT) and **perspectives from** Police directive, EUROPOL directive, Frontex and legal framework in the EU MS; ← T2.1)

Methodological Approach to T3.1

Phase 1: Desk Review

The first phase of T3.1 concerns the development of a desk review, which entails the collection of relevant materials by CENTRIC as well as by all participating partners. Therefore, we ask:

- Partners to identify **10 relevant** guidelines, reports, or papers on FCT data sharing from previous/ongoing projects, organisations, academic sources: *national, EU, group-specific* **(approx. 130+ sources)** – we will provide definition for 'relevant' in template.

- Partners to identify relevant practices, resources, stakeholders, tools, data sharing methods, policies, barriers in the literature. CENTRIC will provide a Document Summary Review Template that partners could use to extract relevant data easily and expediently.

- **Time frame requested**: 2018 onwards (after the introduction) of GDPR. Older documents accepted if they provide context and/or are generally relevant/not impacted by GDPR.

- **Domain**: FCT research; also documents from research domains with similar complexity of data production, data sharing etc (e.g., health sector, military) as some of these might have relevance for RDE.

- **Document types needed on:** Existing and emerging data and EU strategies, infrastructure, tools, methods, resources and training resources, policies, whitepapers, industry reports, academic articles, and books.

- Consolidation of the knowledge to describe landscape and identify gaps (CENTRIC).

Document Summary Review Template: An Overview

- CENTRIC created a Document Summary Review Template to assist partners in analysing effectively the key information needed to contribute towards the deliverable T3.1 Consensus report.
- Information requested by the template includes meta data such as document type, name of publishing organisation, year/author etc.
- The Document Summary Review Template requires partners to input summary information by posing exploratory questions such as:
  - **Who?** Type of partners the document refers to.
  - **Why?** Purpose of data sharing
  - **What?** Size of datasets
  - **How?** Duration of data storage

- **Enablers and barriers?** Good practices and barriers to acceptance of RDE
- Questions on guiding standards and policies
- Risk impact assessment
- Further comment section for additional input by partners.

Phase 2: Expert Consultation

Selection of experts:

- Experts can be from outside LAGO or within LAGO network (consortium, own organisation).
- Each partner was asked to consult 4 experts (or more if possible) as combination of (a) LEA's,
- (b) industry, (c) researchers, (d) policymakers with national or international focus
- (Target number: 52+ experts).

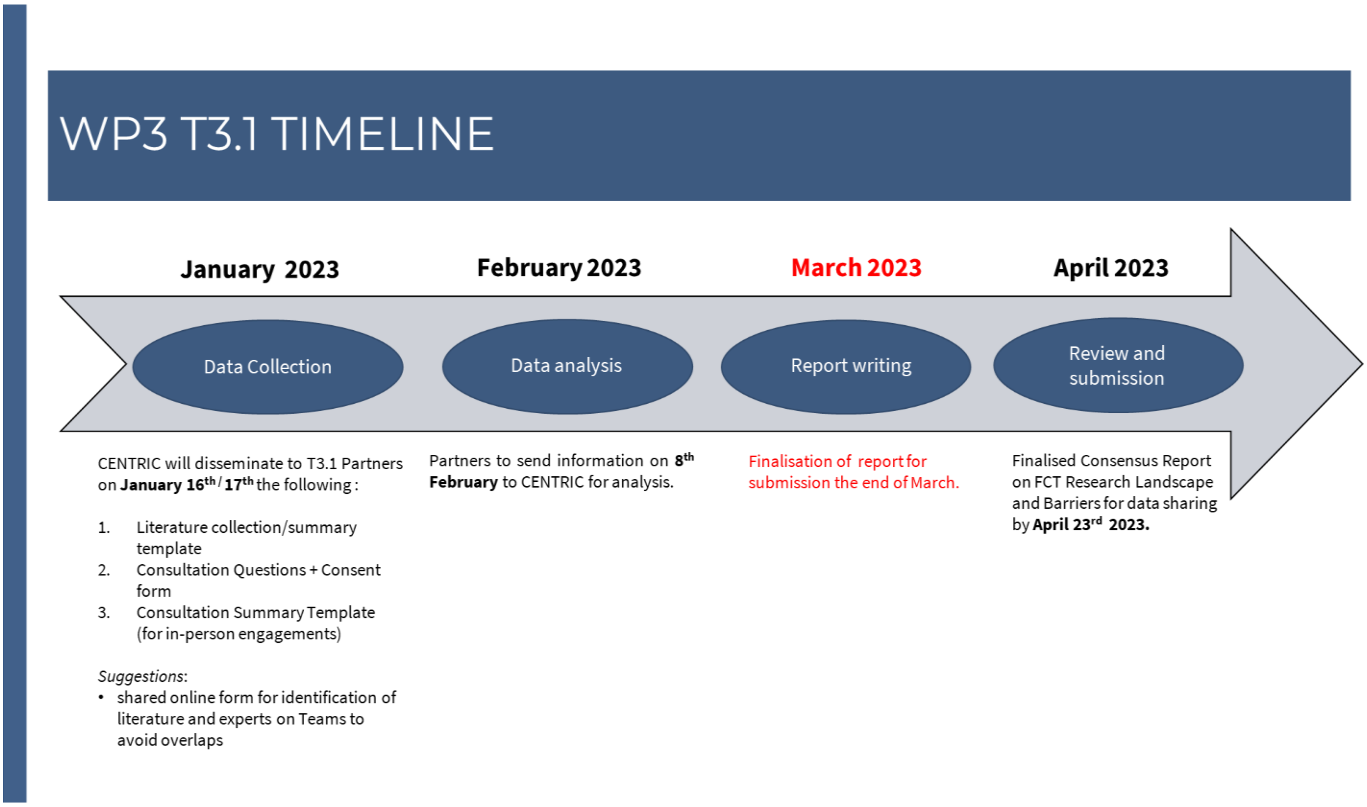
Approach to execution of Phase 2 research activity:

- Provide the expert consultation and consent template for partners to send to identified experts in their network.

Phase 3: Analysis and write up of Consensus Report

- The final stage of the T3.1 will encompass compiling and reviewing submitted document summary reports from partners and analysis of completed written expert consultation forms by CENTRIC. These will be analysed for inclusion and interpretation in the final report. The first draft of the report will be made available for internal reviewers (ICCS and CEA) on the 9th of April for comment. Once it has been reviewed internally it will be forwarded to the Security Advisory Board for the 14th of April for their review. The final report should be made publicly available on the 23rd of April.

WP3 T3.1 Timeline and deadlines



WP3 T3.1 TIMELINE

CENTRIC will disseminate to T3.1 Partners on **January 16th / 17th** the following:

1. Literature collection/summary template
2. Consultation Questions + Consent form
3. Consultation Summary Template (for in-person engagements)

Suggestions:

- shared online form for identification of literature and experts on Teams to avoid overlaps

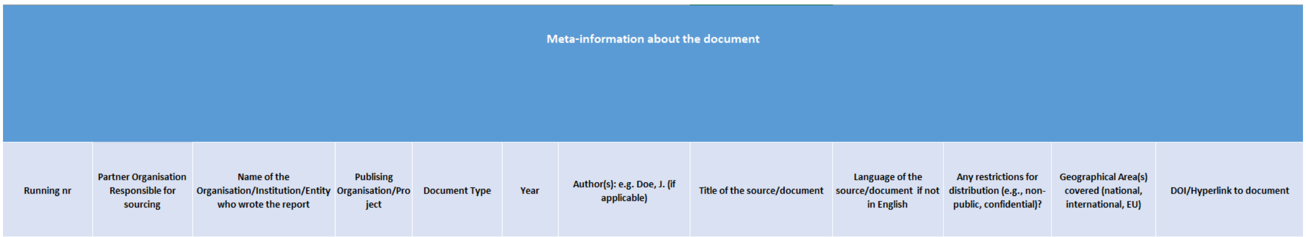
Partners to send information on **8th February** to CENTRIC for analysis.

Finalisation of report for submission the end of March.

Finalised Consensus Report on FCT Research Landscape and Barriers for data sharing by **April 23rd 2023.**

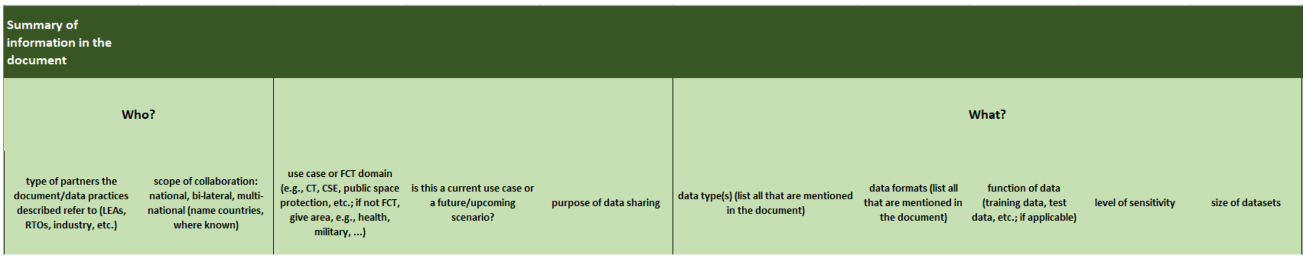
11.2 Appendix B – Document Summaries Template

The following Microsoft Excel™ template was used by the partnering organisations upon recommending literature and documentation for the current report.



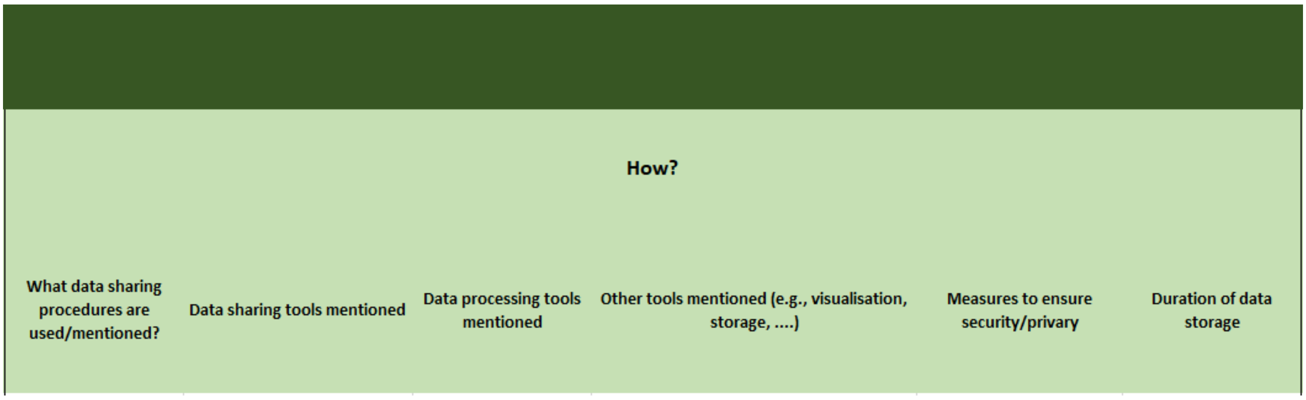
Meta-information about the document											
Running nr	Partner Organisation Responsible for sourcing	Name of the Organisation/Institution/Entity who wrote the report	Publising Organisation/Project	Document Type	Year	Author(s): e.g. Doe, J. (if applicable)	Title of the source/document	Language of the source/document if not in English	Any restrictions for distribution (e.g., non-public, confidential)?	Geographical Area(s) covered (national, international, EU)	DOI/Hyperlink to document

Figure 7: Document summary template: Meta-information about the document



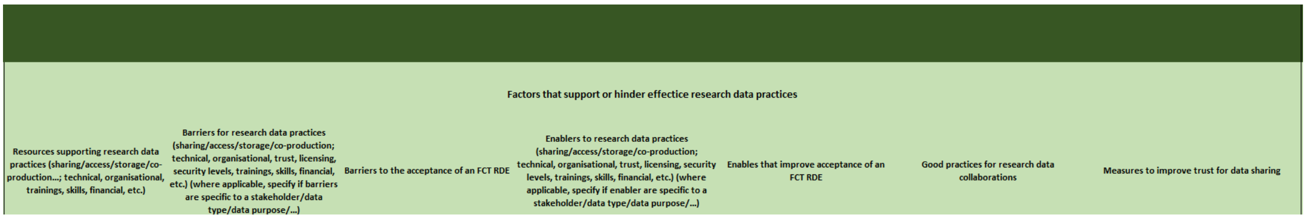
Summary of information in the document									
Who?					What?				
type of partners the document/data practices described refer to (LEAs, RTOs, industry, etc.)	scope of collaboration: national, bi-lateral, multi-national (name countries, where known)	use case or FCT domain (e.g., CT, CSE, public space protection, etc.; if not FCT, give area, e.g., health, military, ...)	is this a current use case or a future/upcoming scenario?	purpose of data sharing	data type(s) (list all that are mentioned in the document)	data formats (list all that are mentioned in the document)	function of data (training data, test data, etc.; if applicable)	level of sensitivity	size of datasets

Figure 8: Document summary template, Summary of information in the document: Who and What



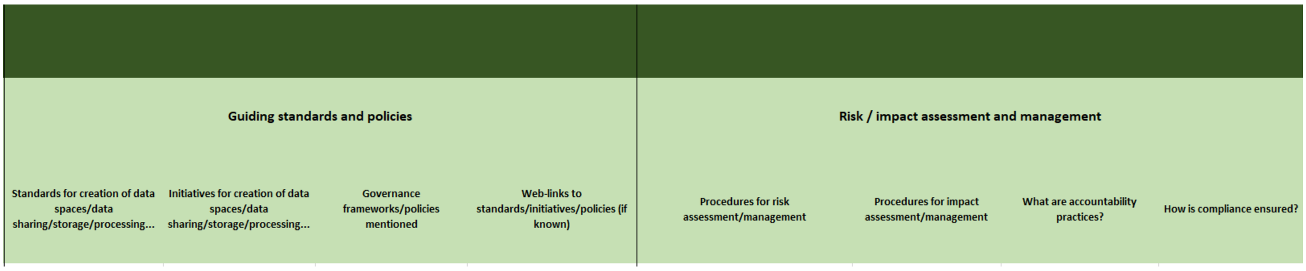
How?					
What data sharing procedures are used/mentioned?	Data sharing tools mentioned	Data processing tools mentioned	Other tools mentioned (e.g., visualisation, storage, ….)	Measures to ensure security/privary	Duration of data storage

Figure 9: Document summary template, Summary of information in the document: How



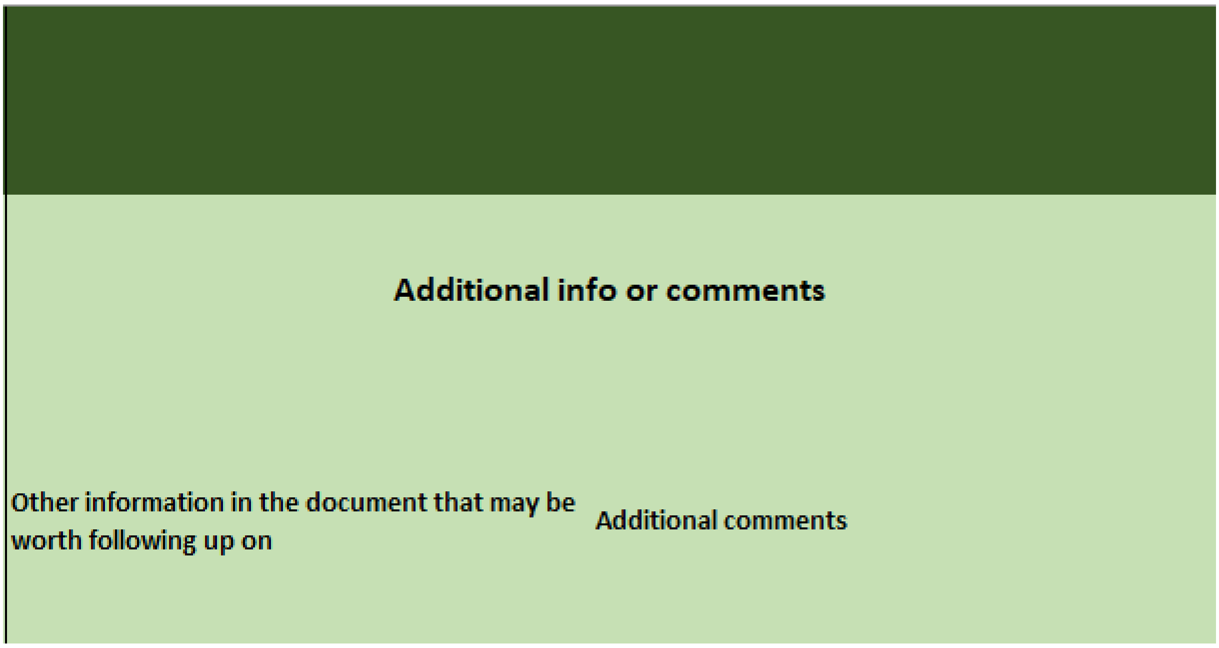
Factors that support or hinder effectice research data practices						
Resources supporting research data practices (sharing/access/storage/co-production...; technical, organisational, trainings, skills, financial, etc.)	Barriers for research data practices (sharing/access/storage/co-production; technical, organisational, trust, licensing, security levels, trainings, skills, financial, etc.) (where applicable, specify if barriers are specific to a stakeholder/data type/data purpose/...)	Barriers to the acceptance of an FCT RDE	Enablers to research data practices (sharing/access/storage/co-production; technical, organisational, trust, licensing, security levels, trainings, skills, financial, etc.) (where applicable, specify if enabler are specific to a stakeholder/data type/data purpose/...)	Enables that improve acceptance of an FCT RDE	Good practices for research data collaborations	Measures to improve trust for data sharing

Figure 10: Document summary template, Summary of information in the document: Factors that support or hinder effective research data practices



Guiding standards and policies				Risk / impact assessment and management			
Standards for creation of data spaces/data sharing/storage/processing...	Initiatives for creation of data spaces/data sharing/storage/processing...	Governance frameworks/policies mentioned	Web-links to standards/initiatives/policies (if known)	Procedures for risk assessment/management	Procedures for impact assessment/management	What are accountability practices?	How is compliance ensured?

Figure 11: Document summary template, Summary of information in the document: Guiding standards and policies & Risk/Impact assessment and management



Additional info or comments	
Other information in the document that may be worth following up on	Additional comments

Figure 12: Document summary template, Summary of information in the document: Additional l info

11.3 Appendix C – Written Expert Consultation Questions

Section 1: Current State of Practice

1. **Current/Existing Initiatives and Data Strategies**

Q1. Are you aware of existing initiatives for the creation of data spaces inside or outside the EU that may be relevant for a Research Data Ecosystem (RDE) in the FCT domain? They can also be in other domains such as health or other areas with similar or lower levels of complexity.

1. If yes, please name these initiatives below.

2. Which elements in the above-mentioned initiatives could LAGO apply to its RDE?

Q2. Are you aware of Data Strategies either from research projects or on national/EU level that are relevant for the FCT domain? Please name all you are aware of.

Q3. Which aspects of the Data Strategies should LAGO apply for its RDE?

2. Current Usage Scenarios

Q1. For which current FCT application domain(s) or usage situation(s) do you consider an RDE to be (a) particularly effective or (b) particularly ineffective? (examples for FCT application domains: counter-terrorism, CSE, Cyber Crime, Terrorism, Firearms and other illegal trafficking, public space protection; examples for usage situations: AI tool development, face recognition, text analysis, video analysis, data creation, data quality assessment and verification tools)

1. particularly effective:

2. particularly ineffective:

Q2. Are there future FTC domains or usage scenarios, the LAGO RDE should support or consider?

3. Current/Existing Standards and Policies

Q1. Which groups/organisations need to be included into an FCT-specific Research Data Ecosystem (RDE)? (e.g., LEAs, specialist functions with LEAs, private companies developing AI solutions, ministries of the interior, data providers, legal experts, …). Please name all groups that you think are necessary for a successful RDE.

Q2. Should any groups/organisations be excluded? If so, why?

Q3. Do the parties you identify have similar or different frameworks, standards, strategies in how they share/co-create and manage data? Please explain.

Q4. Are you aware of standards or policies which regulate data practices that you would consider good practice/gold standards? These can be project specific, national, or international, and in the FCT domain or stem from other domains of similar complexity. (Please note: We are interest in standards and policies outside legal and ethical discussions, as these are captures in a separate effort.)

(d) Data Types and Sharing in the RDE

Q1. What data types are relevant in FCT research?

Q2. Which are the most important data types that LAGO RDE needs to accommodate? Please list at least 5 data types in order of importance.

1.

2.

3.

4.

5.

Q3. Are there requirements for minimal/maximal data sizes for data access, sharing, storage in your project(s)/organisation?

Q5. How is the FCT research data stored in your project(s)/organisation?

Q6. How is the security of the data maintained in your project(s)/organisation?

SECTION 2: Tools, Resources, Barriers and Enablers

1. Tools to support FCT Research

Q1. What tools do you use or are you aware of for supporting effective (a) data access, (b) sharing, (c) co-creation in FCT research?

1. Data access:
2. Data sharing:
3. Data creation/co-creation:
4. Data acquisition:
5. Data anonymisation/pseudonymisation:
6. Data visualisation:
7. Others:

Q2. Are there special requirements that tools need to fulfil to support research in the FCT domain?

2. Resources for FCT Research

Q1. What resources are required for successful FCT research? List at least 5 you consider the most relevant (consider resources addressing disparate aspects such as technical, organisational, governance, licensing, financial, ownership, security levels, policy, training/skills, ...)

1.

2.

3.

4.

5.

3. Most important barriers and enablers

Q1. What are the 5 most important barriers to FCT data sharing? Please rank them in order of importance and provide a short description of the issue/s.

1.

2.

3.

4.

5.

Q2. What are the 5 most important enablers of FCT data sharing? Please rank them in order of importance and provide a short description.

1.

2.

3.

4.

5.

Q3. What are possible barriers and enablers for the adoption of an FCT-specific RDE (if different from the ones you listed above)?

1. barriers for RDE adoption:

2. enablers for RDE adoption:

(b) Trust in FCT research

Q1. In your experience, how much of an issue is trust between partners for data sharing/co-creation in FCT research?

Q2. What are potential drivers of the trust issues you encountered?

Q3. Are there stakeholders that regularly show particularly high or low willingness to share data?

1. If yes, which ones? Please specify if willingness is high or low in the referred instance.

2. Is (un)willingness related to specific datatypes or usage scenarios?

SECTION 3: Good Practices including risk assessment and management

1. Risk assessment and management

Q1. What do you consider the main risks with respect to data creation/sharing in FCT research?

Q2. What risk management approaches or solutions should LAGO adopt for its RDE in the FCT domain?

Q3. What are the most important risks that may jeopardise the success of an RDE in the FCT domain?

2. Good Procedures and Practices

Q1. Can you describe core mechanisms used for data access, transfer, storage, or co-creation in your project(s)/organisation?

Q2. What do you do yourself to ensure effective handling of data in FCT research?

Q3. Give an example from your own experience that you consider a particularly bad data practice in FCT research. Who was involved, what happened and why do you consider it a bad practice? (**Please do not provide personal details of people or organisations.)**

Q4. Give an example from your own experience that you consider a particularly good data practice in FCT research. Who was involved, what happened, and why do you consider it a good practice? **(Please do not provide personal details of people or organisations.)**

SECTION 4: Mapping Future Requirements

1. Opportunities and Gaps for future research and development

Please identity from your own experience **5 main opportunities/gaps** for future research and development in FCT research and data sharing.

Please rank your entries from 1 to 5; 1 being the most important opportunity/gap.

1.

2.

3.

4.

5.

Are there any other aspects/insights you want to share with us?

Contacts for additional consultations

Are you aware of experts or of ongoing or past R&D projects in the FCT domain that we should consult for further information?

Thank you for your support and participation.

11.4 Appendix D – Expert Consultation Consent Form

LAGO

Expert consultation on requirements for a secure and trusted data infrastructure in the FCT domain

Thank you for agreeing to participate in the expert consultation.

Context to the consultation: The consultation is conducted as part of the multi-national project LAGO ("Lessen data access and governance obstacles") funded under the Horizon 2020 scheme (grant nr 101073951): for more information see project website: lago-europe.eu). The ***LAGO project aims to deliver the foundation for a trusted EU FCT Research Data Ecosystem (RDE)*** to address the so-called "Data Issue" in the FCT research landscape, i.e., the lack of domain specific data in sufficient quality and quantity to enable appropriate training and testing of the developed methods, platforms and tools. For this purpose, LAGO will ***develop an evidence-based and validated multi-actor Reference Architecture*** for FCT actors to deposit, share and co-create data and tools for FCT research purposes based on common rules, protocols, standards and instruments in a trusted and secure environment.

Your participation: You have been invited to participate because of your knowledge, experience and expertise relating to current practices, resources, tools, gaps, barriers, and enablers of FCT data sharing. Your input will be instrumental in shaping the Research Data Ecosystem that LAGO will produce. Please note: You are invited in your capacity as expert. hence, your views do not (need to) represent the organisation for which you work.

- **What will you be required to do?** You are asked to provide written input to a list of question (to be found in this document after the consent) about your insights into current practices, resources, tools, gaps, barriers, and enablers of FCT data practices. This consultation should take approx. 40 minutes to complete.
- **Will any of my personal information be exposed as a result of the study?** All data will be pseudonymised by removing all identifying information (names, email addresses, etc.). Data will only be referred to with a running number that does not allow linkages to individuals. All further processing will be conducted on this pseudonymous data and therefore cannot be traced back to you as a participant.
- **What happens to the information once the study is complete?** The information will be summarized in fully anonymised form in confidential reports to the European Union and in academic publications. Raw data will be kept for research purposes for ten years after publication in line with recommendations for academic data verification and accountability.
- **Who will have access to the information?** Access to the pseudonymised data is limited to the research personnel in the LAGO project stored on a dedicated, secured system.
- **Are there any risks involved from participating?** We do not foresee any risks from your participation, but if you are concerned you can contact the principal investigator (contact details see below).
- **What if you decide to no longer take part?** You can stop the consultation at any time without negative consequences. You further have the right to withdraw your data up to 14 days after the consultation. To do so, email the principal investigator for this task (contact details see below).

If you have any questions or concerns about the study, please feel free to contact the principal investigator **Prof Saskia Bayerl at p.s.bayerl@shu.ac.uk.**

Legal basis for the studies. The University undertakes research as part of its function for the community under its legal status. In addition, all University research is reviewed to ensure that participants are treated appropriately, and their rights are respected. This study was approved by UREC with Converis number ER51028819. A full statement of your rights can be found at https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research.

<table>
<tr><td><b>You should contact the Data Protection Officer if:</b>
• you have a query about how your data is used by the University
• you would like to complain about how the University has used your data
• DPO@shu.ac.uk</td><td><b>You should contact the Head of Research Ethics (Dr Mayur Ranchordas) if:</b>
• you have concerns with how the research was undertaken or how you were treated
hscmr@exchange.shu.ac.uk</td></tr>
<tr><td colspan="2">Postal address: Sheffield Hallam University, Howard Street, Sheffield S1 1WBT Telephone: 0114 225 5555</td></tr>
</table>

Informed consent form

Before you start with the consultation, please read the below questions carefully and tick the response that applies.

	YES	NO
1. I have been informed about the LAGO project and the objectives of this research.	□	□
2. I am aware of the context of this study and how my answers will be used.	□	□
3. I understand I have the right to withdraw from the research at any time without providing a reason.	□	□
4. I consent to the information collected in this study to be included, once anonymised, as part of confidential reports to the EU and academic publications.	□	□
5. I understand that I am free to withdraw from the study at any time, without giving a reason for my withdrawal or to decline to answer any particular question without any consequences.	□	□
6. I wish to participate in the study under the conditions set out above.	□	□

Please print, sign, and return together with the filled-out consultation by email.

LAGO Participant Consent

Name: ______________________________ Date: _________________________

Signature: __________________________

11.5 Appendix E – Suggested Literature

Table 18: Recommended by the partners documentation.

Running nr	Publishing Organisation/Project	Document Type	Year	Author(s): e.g. Doe, J. (if applicable)	Title of the source/document	DOI/Hyperlink to document
1	HM Government	Strategy	2021 Sept	Ministry of Defence	Data Strategy for Defence: Delivering the Defence Data Framework and exploiting the power of data	Link
2	HM Government	Strategy	2021 April	Ministry of Defence	Digital Strategy for Defence: Delivering the Digital Backbone and unleashing the power of Defence's data	Link
3	HM Government	Strategy	2020	Ministry of Defence	Defence Data Management Strategy	Link
4	UK Health Data Research Alliance	Strategy	2020		Trusted research Environments (TRE) A strategy to build trust and meet changing health data science needs	Link
5	UK Health Data Research Alliance	Review	2021		Data Research Infrastructure Landscape: A review of the UK data research infrastructure	Link
6	The Lancet Regional Health - Europe	Academic article	2022	Tacconelli, E., Gorska, A., Carrara, E., Davis, R. J., Bonten, M., et al.	Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response	Link
7	Genome Medicine	Academic article	2020	Aarestrup, F. M., Albeyatti, A., Armitage, W. J., Auffray, C., Augello, L., Balling, R., et al.	Towards a European health research and innovation cloud (HRIC)	Link
8	AP4AI	Report	2022	Gibson, S. Heyes, A. Lyle, A. Raven, F. Sampson	Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain	Link
9	Criminal Justice Policy Review	Article	2022	Pickering, J., Fox, A.	Enabling Collaboration and Communication Across Law Enforcement	Link

					Jurisdictions: Data Sharing in a Multiagency Partnership	
10	Medical Writing	Article	2019	Thomas, K., Paarlberg, R.	International Committee of Medical Journal Editors’ requirements for sharing individual participant data from interventional clinical trials	Link
11	Ministry of Defence, Republic of Bulgaria	Policy	2018	Karakachanov, Krasimir	Политика за защита на лиюните данни в министерството на отбраната, структурите на пряко подюинение на министъра на отбраната и българската армия "Policy for personal data protection in the Ministry of Defence and the Bulgarian Army"	Link
12	Privacy Shield	Policy - current initiative			EU-U.S. Privacy Shield Framework Principles issued by the U.S. Department of Commerce	Link
13	Tech Against Terrorism	Report	2021		Transparency Report: Terrorist Content Analytics Platform	Link
14	Nature	Academic Paper	2021	Warnat-Herresthal, S., Schultze, H., Shastry, K.L. et al.	Swarm Learning for decentralized and confidential clinical machine learning	Link
15	Computer Communications	Academic Paper	2019	Domingo-Ferrer, Josep & Farràs, Oriol & González, Jordi & Sánchez, David.	Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges	DOI
16	Research Gate	Academic Paper	2018	Zheng, Zibin & Xie, Shaoan & Dai, Hong-Ning & Chen, Xiangping & Wang, Huaimin.	Blockchain challenges and opportunities: A survey	Link
17	Arxiv	Academic Paper	2019	Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong	Federated Machine Learning: Concept and Applications	Link

18	Ministry of Administrative Reconstruction	Ministry Regulation	2022	Editorial group are officials of the Transparency and Open Government Department of the Directorate for e-Governance of the Directorate-General for Public Organisations of the Ministry of Administrative Reconstruction, coordinated by the National Representative in the OGP	4th National Action Plan on Open Government 2019-2021 (May 2019) Updated Version - Including Addendum with Additional Commitments incorporated through the OGP Greece Ideathon (Dec 2020)	Link
19	European Commission	Act/Regulation	2022	Eu	Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance)	Link
20	International Data Spaces Association	Existing Initiatives/Standards	2021	Prof. Dr. Boris Otto, Fraunhofer ISST Alina Rubina, DE-CIX Management GmbH Andreas Eitel, Fraunhofer IESE Andreas Teuscher, SICK AG Anna Maria Schleimer, Fraunhofer ISST Dr. Christoph Lange, Fraunhofer FIT Dr.-Ing. Dominik Stingl, DE-CIX Management GmbH Evgueni Loukipoudis, DTS Gerd Brost, Fraunhofer AISEC Gernot Böge, FIWARE Foundation e.V. Heinrich Pettenpohl, Fraunhofer ISST Jörg Langkau, nicos AG Joshua Gelhaar,	GAIA-X and IDS	Link

				Fraunhofer ISST Koki Mitani, NTT Corporation Marius Hupperz, Fraunhofer ISST Monika Huber, Fraunhofer AISEC Nils Jahnke, Fraunhofer ISST Robin Brandstädter, Fraunhofer IESE Sascha Wessel, Fraunhofer AISEC Sebastian Bader, Fraunhofer IAIS		
21		Existing Initiatives/Standards			CATENA-X	Link
22	Springer	Existing Initiatives/Standards	2022		Common European Data Spaces: Challenges and Opportunities	Link
23					FRAMEWORK FOR DATA SHARING IN PRACTICE	Link
24	IEEE MultiMedia	Research Article	2020	Bogdan Ionescu; Marian Ghenescu; Florin Răstoceanu; Răzvan Roman; Marian Buric	Artificial Intelligence Fights Crime and Terrorism at a New Level	Link
25	Perspectives on Terrorism, JSTOR	Dataset Collection	2021	Neil Bowie	40 Terrorism Databases and Data Sets: A New Inventory	Link
26	Frontiers in Psychology	Research Article	2022	Jens F. Binder, Jonathan Kenyon	Terrorism and the internet: How dangerous is online radicalization?	Link
27	European Parliamentary Research Service (EPRS)	Report	2022	Carmen-Cristina Cîrlig	Strengthening Europol's mandate	Link
28	European Commission	Proposal for a regulation	2020		Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance(Data Governance Act)	Link
29	European Parliament	Report	2022	Udo Bux / Mariusz Maciejewski	Personal data protection	Link
30	Hellenic Data Protection Authority	Directives	2022	Hellenic Data Protection Authority	Επεξεργασία δεδομένων υγείας	Link

31		Final Report	2020		Study on the practice of direct exchanges of personal data between Europol and private parties	Link
32		Handbook	2017		Handbook on security of Personal Data Processing	Link
33	Publications Office of the European Union, 2019	Science for Policy report	2019	Antofie Tiberiu-Eugen, European Commission, Joint Research Centre (JRC), Ispra, Italy Stefano Luoni, GFT Italia S.r.l external service provider of European Commission, Joint Research Centre (JRC), Ispra, Italy Anna Faiella, traineeship program at European Commission, Joint Research Centre (JRC), Ispra, Italy Montserrat Marin Ferrer, European Commission, Joint Research Centre (JRC), Ispra, Italy,	Risk Data Hub – web platform to facilitate management of disaster risks	Link
34	Springer Open Access	eBook	2022	Edward Curry, Simon Scerri, Tuomo Tuikka (Editors)	Data Spaces: Design, Deployment,and Future Directions	Link
35		pdf /framework	2019		TRUSTED DATA SHARING FRAMEWORK	Link
36	Regional Court (Landgericht) Ulm, Germany	Academic Article	2018	Claudia Warken	Classification of Electronic Data for Criminal Law Purposes	Link
37	PIM Protection Information management	pdf/Report	2018		FRAMEWORK FOR DATA SHARING IN PRACTICE	Link
38		pdf/Report	2021		Sharing Data For Impact: Lessons From Data Sharing Initiatives in Asia	Link
39	International Data Spaces Association	Reference Architecture			International Data Space Reference Architecture Model 4.0	Link
40	Data Spaces Support Centre	White paper	2022		Starter Kit for Data Space Designers	Link
41	International Data Spaces Association	Existing Initiatives/Standards	2021	Otto B. et al.	Gaia-X positioning paper	Link

42	European Commission	Policy Paper	2020		White paper on Artificial Intelligence	Link
43	Arxiv	Academic Paper	2021	Li Q. et al.	A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection	Link
44	Arxiv	Academic Paper	2022	T Salazar, M Fernandes, H Araujo, and P H Abreu	FAIR-FATE: Fair Federated Learning with Momentum	Link
45	MDPI	Academic paper	2022	Gosselin, R.; Vieu, L.; Loukil, F.; Benoit, A	Privacy and Security in Federated Learning: A Survey	Link
46		White paper	2016		FAIR principles	Link
47	The Estonian Academy of Security Sciences	Digital book	2020	Puusalu, J.	Suurandmed: olemus ja kasutamise kitsaskohad	Link
48	Tallinn University of Technology	Master's thesis / digital file	2020	Saarmets, M.	Andmelao metaandmete infosüsteemi analüüs ja kavandamine Tallinna Tehnikaülikooli näitel (Analysis and Design of the Metadata Information System for the Data Warehouse on the Example of Tallinn University of Technology)	Link
49	Ministry of Economic Affairs and Communications of Estonia	Strategy	2021	Ministry of Economic Affairs and Communications of Estonia	Estonia´s Digital Agenda 2030	Link
50	Estonian Ministry of Interior	Strategy	2022		ICT Strategy	
51	European Commission	Standard	2022	SEMIC team	The DCAT Application Profile for data portals in Europe (DCAT-AP)	Link
52	Estonian Ministry of Interior	Strategy	2021	Estonian Ministry of the Interior, Estonian Ministry of Justice	Joint action plan for the digitization of criminal proceedings in the administration of the Ministry of Justice and the Ministry of the Interior	