D1.3: Data Management Plan

Grant Agreement ID	101080100	Acronym	FRISCO
Project Title	Fighting teRrorISt Content Online		
Start Date	15/11/2022	Duration	24 Months
Project URL	https://friscoproject.eu		
Contractual due date	14/05/2023	Actual submission date	12/05/2023
Nature	DMP	Dissemination Level	PU
Author(s)	Antonis Koukourikos (NCSR-D)		
Contributor(s)	Vangelis Karkaletsis (NCSR-D), Charalampos Mavrikas (NCSR-D), Pal Boza (TREMAU), Martina Manfredda (DLEARN), Rositsa Dzhekova (VPN)		
Reviewer(s)	Klara Süveges-Heilingbrunner (IVSZ)		

Document Revision History *(including peer reviewing & quality control)*

Version	Date	Changes	Contributor(s)
v0.1	15/03/23	Document creation and ToC	NCSR-D
v0.2	01/04/23	Sections 3-5	NCSR-D
v0.3	24/04/23	Sections 6-7	NCSR-D
v0.4	02/05/23	First complete draft	NCSR-D
v0.5	12/05/23	Internal Review	IVSZ
v1.0	12/05/23	Report finalisation	NCSR-D
v1.0	12/05/23	Deliverable submission	NCSR-D

Executive Summary

The report outlines how data and metadata will be collected, processed or generated in the context of the FRISCO project. It also details the provisions to ensure data integrity, security, privacy and preservation, as well as the preliminary measures and directions to secure the maximum compliance of FRISCO data with the core principles for making them FAIR (Findable, Accessible, Interoperable and Reusable).

In principle, FRISCO is not data-heavy in terms of the volume and variety of data that will be accessed, collected and/or analysed during its activities. Nevertheless, the project relies on the acquisition of a broad spectrum of information and feedback from relevant stakeholders in order to accurately determine the needs of Hosting Service Providers (HSPs), the expectation of Law Enforcement Agencies (LEAs), and – at the next phase – assess the effectiveness and applicability of the designed tools. As such, the data management plan of the project is critical for ensuring that all collected information is handled ethically and in accordance with relevant regulations. Furthermore, the data management plan sets the pathways that will facilitate the generation of meaningful impact by the project's data outputs.

As the project activities progress and evolve, the aspects affecting data management will naturally concretise. Consequently, the Data Management Plan will also be updated to reflect the advancements in understanding of the nature and complexity of data assets and processes. Thus, the DMP will be a living document, regularly updated through the duration of the project.

Table of Contents

List of Figures

No table of figures entries found.

List of Tables

List of Terms & Abbreviations

Abbreviation	Definition
FAIR	Findable, Accessible, Interoperable, Reusable (data)
DMP	Data Management Plan
EC	European Commission
HSP	Hosting Service Provider
LEA	Law Enforcement Agency
TCO	Terrorist Content Online
DOI	Digital Object Identifier
PII	Personally Identifiable Information
EU	European Union
IPR	Intellectual Property Right
CA	Consortium Agreement
GA	Grant Agreement

1   Introduction

1.1   Purpose and Scope

The purpose of deliverable *D1.3, Data Management Plan* is to describe the data management life cycle for the data to be collected, processed and/or generated by the FRISCO project. As part of making research data findable, accessible, interoperable and reusable (FAIR), the project's Data Management Plan (DMP) includes information on the handling of data during & after the end of the project; what data will be collected, processed and/or generated; which methodology & standards will be applied; whether data will be shared/made open access; how data will be curated & preserved (including after the end of the project).

1.2   Approach to Work Package and Relation to other Work Packages and Deliverables

Data management falls under WP1, the work package responsible for the coordination and assessment of all FRISCO activities. As a critical aspect of the work plan, it informs all activities under the different work packages, in the sense that data collection and analysis will inform the direction of the tools to be developed, the training material to be designed and published, as well as the focus, modality and context of the various dissemination and extroversion activities to be included in the respective plan. Furthermore, the acquisition and analysis of feedback from the aforementioned activities is of substantial importance for the project itself and the sustainability and relevancy of its main outcomes.

As a result, the Data Management Plan is bidirectionally linked to all work packages, needing their input to properly defined user needs and expectations, and – on the other hand – setting up the directives and guidelines to ensure the efficient and secure management of all project data. As activities under the different work packages progress, the Data Management Plan will accordingly adapt to new information and findings.

1.3   Methodology and Structure of the Deliverable

The deliverable is structured in accordance with the template and guidelines provided by the EC, and is organized in the following sections: Section 2 provides a data summary addressing issues regarding the purpose of the data collection/generation and its relation to the objectives of the project, the types and formats of data the project will generate/collect, the origin of data, the expected size of data, and the data utility (i.e., to whom might it be useful). Section 3 reports on the measures and directions to be adopted to ensure the compliance of FRISCO with FAIR data principles. Section 4 summarizes the allocation of resources and a preliminary cost coverage plan in the context of the project, in order to serve the aforementioned measures. Section 5 discusses the main concerns regarding data security and privacy and the proposed approaches to face them, while Section 6 discusses ethical and legal aspects related to the data involved in the project.

2   Data Summary

FRISCO objectives are explicitly linked to the analysis of the current positioning, the established methods and processes, and the knowledge and technology gaps faced by the targeted HSPs in order for them to be compliant with the TCO regulation. Consequently, and to develop its methodologies and tools, FRISCO will extensively contact micro and small HSP representatives to elicit requirements, communicate its methods and test the tools to support them in their work.

To this extent, certain data will be collected during the project, at its initial mapping and requirements elicitation phase, but also throughout its duration for feedback collection and refinement of the tools and instruments produced in the context of WPs 2, 3 and 4.

At the present stage, we identify the following main types of data that will be collected, analysed, or produced by FRISCO.

Table 1: Main types of data in FRISCO

No.	Data Type	Description
1	Questionnaire responses	Documents including the responses of different stakeholders (HSP representatives, LEA representatives) to the questionnaire built in the context of WP2 for guiding targeted interviews
2	Online questionnaire responses	Documents including the responses of stakeholders (HSP representatives) to the online questionnaire built in the context of WP2
3	Feedback forms	Documents including the responses on the questions aiming to collect feedback on the applicability and usage of the FRISCO tools, to be used during the tool assessment phase of the project
4	Sample data for identifying or explaining TCO	Data snippets that will be used for exemplifying TCO both for machines (model training) and human agents (providing examples to relevant stakeholders), along with any relevant metadata
5	Publications	Scientific or general-audience publications to be produced in the context of the technical and dissemination activities of FRISCO
6	Training material	Content of different modalities (documents, website areas, presentation, multimedia content) produced in the context of WP4
6	Source code	Source code of the digital tools that will be developed by the project

As further detailed in the following sections, all data collection will comply with regulations pertaining to the provision of consent, anonymisation of processing, secure management and any other measure that ensures compliance with rules and ethics directives, in conjunction with the adoption of processes that will ensure the FAIRness - to the extent possible – of all data assets.

3   FAIR Data

3.1   Making data findable, including provenance for metadata

For the fair access of the project data a Digital Object Identifier (DOI) will be requested for each artefact. In more details, DOIs from Crossref[1] will be used for scientific publications, while DOIs from DataCite[2] will be pursued for labelling each dataset of the project. In addition, appropriate metadata records for each output of the project will be created and stored in the relevant data directory, covering the specification of the relevant service as well as using elements from appropriate ontologies and taxonomies, such as D-CAT, DC-TERMS, etc. Amongst other fields, each metadata record will have a set of keywords that will make searches easier for external parties. Each data source will be provided with a specific name that is composed by different parts/elements, containing spatiotemporal coverage, methodology, data types and formats where relevant, naming conventions and any other facet that contributes to the findability (and reusability) of the data source.

3.2   Making data openly accessible

Several datasets that will be used as part of the project will be offered by third-party providers, as summarised in Table 1, with some of them being confidential, proprietary and of commercial or social sensitivity. In the cases where private data are processed and aggregated (e.g., as part of a model, or functionality of a component) permission will be requested by the provider prior to making the altered data publicly available after processing for PII.

In reference to the nature of user and contact data involved, some of the results that will be generated by each project phase will be restricted to authorised users, while other results will be publicly available. As per our Ethics commitment, data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU.

Datasets characterised as openly accessible will be published in the Zenodo[3] repository.

Furthermore, and in accordance with EC and Horizon Europe regulations and the Grant Agreement, scientific and popular publications produced in FRISCO will be made available under Open Access models, and accessible via the project and partners portals. Upon decision and occurence, certain publications may be published under Gold Open Access and consequently available via the publisher's website.

In any case, pre-prints of publications will be submitted to relevant archiving services such as arXiv[4].

---

[1] https://www.crossref.org

[2] https://datacite.org

[3] https://zenodo.org

[4] https://arxiv.org

In principle, the source code for software developed in the context of FRISCO will be openly available under appropriate licences, to be selected per case. Publishable source code will be accessible via a central repository hosted in popular relevant services like GitHub or GitLab, with the possibility that the consortium uses repositories accessible only to FRISCO partners for development and testing purposes. Proprietary code will remain the Intellectual Property of its provider/creator, with the ability to expose certain snippets or components as open-source.

As the project evolves the DMP will incorporate details on the services and repositories used for distributing, documenting and supporting the source code.

3.3 Making data interoperable

Interoperability in FRISCO refers mainly to the ability to share aggregate analytical data on the status and knowledge of HSPs regarding TCO and the challenges it brings, as well as, annotated data for identifying and classifying TCO.

The challenge will be met primarily via the extensive usage of semantic web technologies and the incorporation of standards and widely used schemas, ontologies and vocabularies as the foundational blocks of the metadata describing FRISCO data assets. Where applicable, data will be made available via standard-based APIs.

Relevant documentation for the way data is represented will be included in all cases and will be made available following the same approach as the project's source code, i.e., via publicly accessible repositories.

3.4 Increase data reuse

If possible, datasets will be licensed under an Open Access licence. However, this will depend on the level of privacy, and the Intellectual Property Right (IPR) involved in the data primarily for those obtained by HSPs. A period of embargo will only be necessary if a data set contains specific IPR or other exploitable results that will justify an embargo. Therefore, the data will be licensed to permit the widest reuse possible when no limitations are identified by the key stakeholders.

The intention of FRISCO is to make as much data and findings as possible reusable for third parties. Restriction will only apply when privacy, IPR, or other exploitations ground are in play. All datasets will be cleared of bad records, with clear naming conventions, and with appropriate metadata.

All data generated and collected in FRISCO will undergo a quality check in order to analyse its individual plausibility and consistency, making sure that others can directly use it to perform assessments and validate the research and analysis carried out by the project.

4 Allocation of Resources

Data Management is a core overarching activity for FRISCO, as data collection and processing inform its technical activities. Structurally, the activity is the focus of Task T1.4, led by NCSR-D. In any case, effort on Data Management will be also dedicated in work packages 2, 3 and 4. Regarding additional costs emerging from the adopted approach for data and knowledge publishing, the following tables provide a high-level summary.

Table 2: Additional data-related costs and proposed coverage approach

Issue	Action
Costs for making non-patented data and code FAIR	Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis. Possible benefits from country-level/institutional deals with publishers will also be considered. Project Website operation: to be determined Data archiving at OPENAIRE: free of charge Copyright licensing with Creative Commons: free of charge
Partner responsibilities	Every partner is responsible for the data they produce. Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s) in compliance with the CA: During the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedures defined in the Grant Agreement.
Long-term preservation	Data preservation of at least 5 years after the project is required by the Grant Agreement. The associated costs for dataset preparation for archiving will be covered by the project itself. Long-term preservation of code and open datasets will be provided, and associated costs covered by a selected disciplinary repository. Proprietary and sensitive datasets will remain the property of the owners.

5 Data Security

5.1 Security scope in the context of FRISCO

FRISCO activities entail the acquisition, processing and generation of data that pose specific requirements in terms of security, mainly stemming from the fact that we will deal with personal and, potentially, sensitive information collected from HSPs and our interaction with individuals and organisations (interviews, questionnaires). Thus, privacy risks are the main security challenge to be met by the project, with the mechanisms mentioned in Section 2 and 3 to be enforced across project activities and for its entire duration.

5.2 Data protection and GDPR compliance

To ensure that risks related to privacy and sensitivity of information are mitigated, all partners involved in the processing of personal data will be required to apply the appropriate data protection measures as outlined in the General Data Protection Regulation. These may include the collection of consent forms by the contacted individuals, the formulation of appropriate privacy policies and the incorporation of secure registration, authentication and authorization mechanisms for the project's platforms, services and data/document repositories.

6   Ethical Aspects

Potential ethical implications applicable to FRISCO will be managed in the context of WP1, as an integral part of the project's management processes. The WP will set out and establish the general framework for an effective ethical management and the rules that the project must comply with.

The relevant quality reports will provide information on the informed consent procedures with regard to personal data processing and the overall ethical framework for maximum research integrity maintenance. In addition, it describes the logic and architecture that covers the recruiting processes and provides relevant consent forms, following an internal GDPR consent assessment in order to identify and flag risks and revert on these accordingly. Its main points are summarised below, with further details and templates for data collection included in the relevant deliverables upon occasion.

The ethical principles driving FRISCO with respect to data collection and use of personal data for every given task of the project take into consideration the following parameters:

- The personal data FRISCO collects and the purpose(s) for which it collects that information;
- How FRISCO uses personal data internally;
- Whether FRISCO shares personal data with external entities, the categories of those entities, and the purposes for such sharing;
- How individuals may obtain access to their personal data

The consent forms to be used throughout FRISCO comprise sections adhering to all the aspects above, with the forms to be used at each case to be configured by the relevant responsible partners and under communication with the coordinator. In any case, consent provision is explicit, and individuals are explicitly requested to do so 'in time', namely always prior to any processing activity thereof. On the accountability side, all consents are stored, signed, and time stamped.

7   Other Issues

At the moment of preparation of the initial version of the Data Management Plan, no further issues were identified. Any issue that may emerge during the course of the project will be reported at updated versions of the DMP, until the conclusion of the project.

8   Conclusions

The present report constitutes the first version of the FRISCO Data Management Plan. Its aim is to specify the basic principles and guidelines that will inform the management and distribution of all different data assets that will be produced or processed in the context of the project. As FRISCO activities progress, changes and specific cases will naturally be encountered. Thus, the data management plan will be updated accordingly to accommodate the new findings, while keeping in any case the core requirements of data openness, integrity and security.

------End of Document------