Common European
Data Spaces and
Robust AI for Transparent
Public Governance

CEDAR

Project acronym: CEDAR

Project full title: Common European Data Spaces and Robust AI for Transparent Public Governance

Call identifier: HORIZON-CL4-2023-DATA-01

Type of action: HORIZON-RIA

Start date: 01/01/2024

End date: 31/12/2026

Grant agreement no: 101135577

D7.4 Data Management Plan

Document description: ***List of datasets used and generated in the project, associated data management aspect.***

Work package: ***WP7***

Author(s): ***Jolanda Modic, Gilda De Marco, Svitlana Stepanenko, Mounir Kellil, Félix Cuadrado, José Miguel Blanco, José María del Álamo, Oleh Melnychuk, Isabela Maria Rosal, Triantafyllos Kouloufakos, Charlotte Somers***

Editor(s): ***Irem Goymen, Dr. Sophia Karagiorgou***

Leading partner: ***UBI***

Participating partner: ***UPM, ENG, KUL, ALBV, DBC***

Version: ***1.0***	**Status:** ***Complete***
Deliverable type: ***DMP — Data Management Plan***	**Dissemination level:** ***PU - Public***
Official submission date: ***30/06/2025***	**Actual submission date:** ***30/06/2025***

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101135577

Disclaimer

This document has been produced in the context of the CEDAR Project. This project is part of the European Union's Horizon Europe research and innovation programme and is, as such, funded by the European Commission. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. All information in this document is provided 'as is' and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability with respect to this document, which is merely representing the authors' view.

This document contains material, which is the copyright of certain CEDAR contractors, and may not be reproduced or copied without permission. All CEDAR consortium partners have agreed to the full publication of this document if not declared "Confidential". The commercial use of any information contained in this document may require a license from the proprietor of that information. The reproduction of this document or of parts of it requires an agreement with the proprietor of that information.

The CEDAR consortium consists of the following partners:

No.	Partner Organisation Name	Partner Organisation Short Name	Country
1	**Centre for Research and Technology Hellas**	CERTH	Greece
2	**Commissariat al Energie Atomique et aux Energies Alternatives**	CEA	France
3	**CENTAI Institute S.p.A.**	CNT	Italy
4	**Fundacion Centro de Technologias de Interaccion Visual y Comunicaciones VICOMTECH**	VICOM	Spain
5	**TREBE Language Technologies S.L.**	TRE	Spain
6	**Brandenburgisches Institut für Gesellschaft und Sicherheit GmbH**	BIGS	Germany
7	**Christian-Albrechts University Kiel**	KIEL	Germany
8	**INSIEL Informatica per il Sistema degli Enti Locali S.p.A.**	INS	Italy
9	**SNEP d.o.o**	SNEP	Slovenia
10	**YouControl LTD**	YC	Ukraine
11	**Artellence**	ART	Ukraine
12	**Institute for Corporative Security Studies, Ljubljana**	ICS	Slovenia
13	**Engineering – Ingegneria Informatica S.p.A.**	ENG	Italy
14	**Universidad Politécnica de Madrid**	UPM	Spain
15	**Ubitech LTD**	UBI	Cyprus
16	**Netcompany-Intrasoft S.A.**	NCI	Luxembourg
17	**Regione Autonoma Friuli Venezia Giulia**	FVG	Italy
18	**ANCEFVG – Associazione Nazonale Construttori Edili FVG**	ANCE	Italy
19	**Ministry of Interior of the Republic of Slovenia / Slovenian Police**	MNZ	Slovenia
20	**Ministry of Health / Office for Control, Quality, and Investments in Healthcare of the Republic of Slovenia**	MZ	Slovenia
21	**Ministry of Digital Transformation of the Republic of Slovenia**	MDP	Slovenia
22	**Celje General Hospital**	SBC	Slovenia
23	**Transparency International Deutschland e.V.**	TI-D	Germany
24	**Katholieke Universiteit Leuven**	KUL	Belgium
25	**Arthur's Legal B.V.**	ALBV	Netherlands
26	**DBC Diadikasia**	DBC	Greece
27	**The Lisbon Council for Economic Competitiveness and Social Renewal asbl**	LC	Belgium
28	**SK Security LLC**	SKS	Ukraine
29	**Fund for Research of War, Conflicts, Support of Society and Security Development – Safe Ukraine 2030**	SU	Ukraine
30	**ARPA Agenzia Regionale per la Protezione dell' Ambiente del Friuli Venezia Giulia**	ARPA	Italy

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by

0.1	9/5/2025	ToC regarding the HE DMP Template	UBI, Sophia Karagiorgou
0.2	13/5/2025	ICS Contribution to Section 2.1.2	Jolanda Modic
0.3	14/5/2025	INS contribution to Section 2.1.1	Gilda De Marco
0.4	16/5/2025	YC contribution to Section 2.1.3, CEA contribution to Section 6	Svitlana Stepanenko, Mounir Kellil
0.5	19/5/2025	UPM contribution to Section 3 and 4	Félix Cuadrado, José Miguel Blanco, José María del Álamo
0.6	20/5/2025	ART contribution to Section 2.2.3	Oleh Melnychuk
0.7	22/05/2025	KUL Contribution to Section 7	Isabela Maria Rosal, Triantafyllos Kouloufakos, Charlotte Somers
0.8	22/05/2025	CEA contribution to sections 2.2.3 and 4	Mounir Kellil
0.9	13/06/2025	UBI; deliverable finalised and sent for review	Sophia Karagiorgou
1.0	25/06/2025	UBI; addressed reviewers comments and sent to Coordinator for submission	Irem Goymen, Sophia Karagiorgou
2.0	27/06/2025	CERTH; final version and submission	Thodoris Semertzidis, Mariana Minopoulou

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise.

Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

List of Tables

List of Terms and Abbreviations

AI	Artificial Intelligence
CI/CD	Continuous Integration and Continuous Deployment
DAPS	Dynamic Attribute Provisioning Service
DMP	Data Management Plan
DOCX	Document Extensible Markup Language
ECHR	European Convention on Human Rights
EDC	Eclipse Dataspace Connector
FAIR	Findable, Interoperable and Reusable
GDPR	General Data Protection Regulation
HEU	Horizon Europe
HTTPS	Hypertext Transfer Protocol Secure
JSON	JavaScript Object Notation
MVD	Minimum Viable Dataspace
PDF	Portable Document Format
SFTP	Secure File Transfer Protocol
KPK	Slovenian Anti-Corruption Agency
OPSI	Slovenian open data portal
TXT	Text File Document
YAML	Yet Another Markup Language

Executive Summary

The Data Management Plan (DMP) of the CEDAR Project presents a detailed data management approach for data recording, along with the measures and the policies which are applicable and aligned with the Horizon Europe (HEU) requirements for open and private data handling. This revised version of the deliverable, succeeding its predecessor deliverable D7.2 Data Management Plan delivered at M06, documents the undertakings pursued by the CEDAR Consortium to making data Findable, Interoperable, and Reusable (FAIR) to the broadest possible audience.

The DMP details the manner in which data is made accessible, collected, stored, and disseminated, with a specification of secure data exchange and interoperability standards. It addresses data security and ethics, including compliance with the General Data Protection Regulation (GDPR) and EU regulations. The CEDAR Consortium has established a DMP Register which functions as a living document, tracking data specifications, research, and software artifacts, ensuring alignment with open data principles.

1 Introduction

This deliverable provides a comprehensive revision of the CEDAR Data Management Plan (DMP), as collectively contributed by all Consortium Partners and registered by M18 of the project. In alignment with the European Commission's Guidelines for the Horizon Europe (HEU) Programme [1], the status of the DMP has been revisited to incorporate the advancements achieved until M18.

Deliverable D7.4 presents updates on data-related activities concerning the collection, harmonization, synthesis, and processing during the reporting period. Consistent with the monitoring of data evolution within CEDAR, the DMP Register is maintained as an up-to-date record within the project's centralized repository. As of M18, the DMP Register has been monthly updated to reflect the current status of available, collected, generated, and processed data, with the subsequent update scheduled for D7.6 Data Management Plan at M36.

Specifically, in pursuit of the principles aimed at ensuring that research data collected or generated over the course of and subsequent to the project is Findable, Accessible, Interoperable, and Reusable (FAIR), Deliverable D7.4 elaborates on the CEDAR Consortium methodology implemented to fulfil this objective. Accordingly, this deliverable outlines — among other aspects— how the collected, generated, and processed research data has been managed during the reporting period and following the conclusion of the CEDAR Project. Details regarding data enhancements per pilot are provided in conformity with FAIR principles. The activities undertaken are consistent with Internal Data Collection and Enrichment protocols. Concurrently, updates are provided on synthetic data generation within CEDAR and the integration of External Data Collections from open sources. Standards and methodologies for data identification, collection, recording, generation, and sharing within the CEDAR Data Space are elaborated upon. Finally, progress concerning the advancement of research data results is reported. It should be noted that this document is fundamentally based on the related template provided by the European Commission.

1.1 Purpose of the Document

The Data Management Plan describes how the data has been collected, classified, stored and made FAIR within the CEDAR Project. It specifies the type of data that is generated and collected during the project, the serialization and interoperability standards that are used, how the research data results are preserved and what parts of the datasets are made available to be exchanged and shared via the CEDAR Data Space for verification and reuse.

The revised plan is aligned with the progress of the project. The CEDAR DMP addresses the following aspects:

- The overview of the Data Management approach in CEDAR for the 2nd version of the Datasets;
- The methodology adopted to make data FAIR;
- Other research outputs that are closely monitored in DMP;
- Data security; and
- Data ethics.

1.2 Relation with Other Tasks and Deliverables

Task 7.2, encompassing Scientific, Technical, and Data Management in collaboration with the Pilot and Technical Partners, is responsible for the identification of datasets and the formulation of the methodology to be implemented during and subsequent to the project. This is to optimize the generation of synthetic data and facilitate access to and reuse of the data models available for the CEDAR Data Space, in alignment with data protection and open access policies applicable to selected datasets. This task represents an ongoing endeavour throughout the CEDAR Project lifecycle, with reporting conducted via the evolving document of the DMP Registry.

Deliverable D7.4 also encompasses the datasets that have been collected and produced within WP1-WP5, in adherence to the FAIR principles. Moreover, it documents the efforts and activities undertaken by the Consortium to maintain the currency of the data, in synchronization with the progression of the technical activities of the project. Furthermore, D7.4 describes the manner in which data is stored and secured, along with identifying where and by whom, as well as determining the access permissions. In this way, the Data Management Plan aligns with CEDAR's mission of advancing open science and data sharing, as it addresses the entire data management lifecycle, ensuring rigorous documentation, data curation, and data archival for prospective projects.

1.3 Deliverable Structure

The structure of this document is as follows:

Chapter 2 reports the progression of data within the pilot contexts since M06, which forms the foundation for advancing the project's technical activities. This chapter also encompasses the data type and format, its purpose, size, and origin. Any instance of reusing existing data within the project is documented, along with the justification for its reuse. Furthermore, the potential applications of this data beyond the confines of the project are elucidated.

Chapter 3 presents the strategies employed to guarantee the data's Findability, Accessibility, Interoperability, and Reusability (FAIR).

In Chapter 4, a comprehensive account of additional research artefacts, results, and outputs identified or repurposed within the project is provided, including data models, studies, and new materials.

Chapter 5 outlines the current strategies to ensure data security, encompassing its storage and recovery.

Chapter 6 discusses any ethical or legal considerations that may influence data sharing. Further, considerations related to personal data, such as informed consent or long-term preservation, are examined.

Finally, Chapter 7 provides the conclusion to the deliverable.

2 CEDAR Data Management (2^{nd} Version)

2.1 Enhancements in CEDAR Datasets

2.1.1 Pilot 1: Transparent Management of National RRP Funds in Italy

The datasets related with Pilot 1 is led by INS and the activities are performed in the scope of Task 5.1. In relation to the initial version of the DMP concerning the regional procurement data available on the eProcurement platform eAppalti, we have thoroughly analysed all the data generated during the management of a tender to identify significant data for CEDAR analyses, eliminating insignificant, erroneous and redundant data, such as duplicated data due to the transition from one phase to another in the tender lifecycle, and optional data that the operator can enter at their discretion. The table that follows highlight the nature and categories of public tenders used for Pilot 1.

The result of this data analysis is a dataset that allows to represent the main characteristic of public tenders, relevant to carry out the type of analysis identified for DEMO1. We extracted from Regional platform real data of tenders for the years 2023 and 2024.

Table 1. Data provided by INS-FVG for Pilot 1.

Number of datasets identified	1
Partner	INS-FVG
Dataset description	885 real tenders covering the three main categories (works, services and goods) from all internal units of Regional Administration related to years 2023 and 2024. It includes all main attributes of a tender from the creation to the closure of the tender, both in case of awarding and closure without any awarding. Moreover, in some type of tender, the dataset includes data on bidders and their positioning in relation to the tender (such as individual bidders, consortium of bidders)
Related WP/Task	WP5 / Task 5.1
Will you reuse any existing data? If YES, how?	NO
Methodologies for data collection / generation	Database extraction
Data type(s)	Text
Data format(s)	Excel files
Data storage	INSIEL
Expected size of the data	2,5 MB
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns and anomalies.
Data access	Stored on the project SharePoint.
Data privacy	INSIEL anonymized the personal data present in the dataset. The data are available in Sharepoint, and has been shared with CEDAR consortium.

During the reporting period, we have also collected relevant satellite imagery data. Regarding the satellite imagery data, we identified a specific area of Trieste, that is under renovation and where are present several ongoing construction sites, that represent the execution of large public tenders, in terms of complexity and value, managed by regional local authorities.

2.1.2 Pilot 2: Transparent Management of Slovenian Public Healthcare Funds

The datasets related to Pilot 2 are led by ICS and the activities are performed in the scope of Task 5.2. Since the initial version of the DMP, we have expanded and refined our dataset collection and preparation activities. These enhancements aim to support more robust anomaly detection, contextual analysis, and validation of analytical tools under development within the DEMO2. The key enhancements are described below. The datasets described in D7.2 are still in use. The tables that follow highlight the nature and diversity of data used for Pilot 2.

Real-world procurement data from SBC: Building on the initial dataset of tenders and bids, we have now collected and cleaned 100 real samples of low-value procurement cases from the SBC archive, covering medical goods, non-medical goods, and services from the years 2022 and 2023. Each sample includes full documentation, namely tender descriptions, winning bids, and associated purchase orders. Additionally, we have collected all tenders published on the SBC website for years 2022 - 2025.

Table 2. Data provided by SBC for Pilot 2.

Number of datasets identified	1
Partner	SBC
Dataset description	100 real samples of low-value procurement cases from the SBC archive, covering medical goods, non-medical goods, and services from the years 2022 and 2023. Each sample includes full documentation, namely tender descriptions, winning bids, and associated purchase orders.
Related WP/Task	WP5 / Task 5.2
Will you reuse any existing data? If YES, how?	Yes. Analysis of procurement practices, identification of patterns and potential anomalies.
Methodologies for data collection / generation	Generated and collected by SBC as part of the procurement process.
Data type(s)	Text
Data format(s)	Word documents, PDFs, Excel files
Data storage	SBC
Expected size of the data	13 MB
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns and anomalies.
Data access	Stored on the project SharePoint.
Data privacy	SBC private data. Shared with the consortium. To be opened.

Table 3. Data provided by SNEP for Pilot 2.

Number of datasets identified	1
Partner	SNEP
Dataset description	All tenders published on the SBC website for years 2022 – 2025.
Related WP/Task	WP5 / Task 5.2
Will you reuse any existing data? If YES, how?	Yes. Analysis of procurement practices, identification of patterns and potential anomalies.
Methodologies for data collection / generation	Generated by SBC as part of the procurement process, published on their website. SNEP used automated web data extraction tools to download the documents in accordance with the SBC website's terms of use.
Data type(s)	Text
Data format(s)	Word documents, PDF, Excel files.
Data storage	SNEP
Expected size of the data	1 GB
Metadata and standards	No particular metadata or standards have been followed while collecting the data. Metadata exists within files and can be analysed later.
For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns and anomalies.
Data access	Published on SBC web page. Accessible with web scraper / crawler.
Data privacy	Public data.

Financial transactions from ERAR: We have obtained and processed detailed financial transactions of SBC from the ERAR portal, managed by the Slovenian Anti-Corruption Agency (KPK). This dataset spans five years (2019–2024) and allows for cross-analysis of spending patterns against procurement and bidder data, offering deeper insights into potential anomalies in fund allocation.

Table 4. Data obtained by ERAR portal by SNEP for Pilot 2.

Number of datasets identified	1
Partner	SNEP
Dataset description	Financial transactions of SBC as collected by the Slovenian Anti-Corruption Agency (KPK) over five years (2019-2024).
Related WP/Task	WP5 / Task 5.2
Will you reuse any existing data? If YES, how?	Yes. Analysis of procurement practices and business relationships, identification of patterns and potential anomalies.
Methodologies for data collection / generation	Gathered via API, and FTP, CSV2SQL transformer. Still waiting for ERAR web page update. API is currently not working due to site maintenance.
Data type(s)	Text
Data format(s)	JSON, CSV
Data storage	https://erar.si/ https://www.enarocanje.si/#/english For the purposes of CEDAR, data is downloaded and stored at SBC infrastructure.
Expected size of the data	20% of tenders volume equals around 200 MB.
Metadata and standards	No particular metadata or standards have been followed while collecting the data.

For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns, relations, and anomalies.
Data access	Accessible through API on request. (API is in development.)
Data privacy	Public data.

Low-value tenders from the national procurement portal (eJN): A new dataset has been integrated from the national public procurement portal eJN, specifically focusing on low-value tenders for the period 2021–2023. This dataset allows us to compare SBC's procurement practices with those of other institutions, helping to identify outliers and systemic patterns.

Table 5. Data obtained by eJN portal by SNEP for Pilot 2.

Number of datasets identified	1
Partner	SNEP
Dataset description	Low-value tenders for the period 2021–2023 as collected by the national public procurement portal.
Related WP/Task	WP5 / Task 5.2
Will you reuse any existing data? If YES, how?	Yes. Analysis of procurement practices and business relationships, identification of patterns and potential anomalies.
Methodologies for data collection / generation	Gathered via API, and FTP, CSV2SQL transformer. Gathered via scrapping / crawling.
Data type(s)	Text
Data format(s)	CSV
Data storage	https://www.enarocanje.si/#/english For the purposes of CEDAR, data is downloaded and stored at SNEP infrastructure.
Expected size of the data	100 MB
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns, relations, and anomalies.
Data access	Public data.
Data privacy	Public data.

Consumer prices and inflation from the Slovenian open data portal (OPSI): To contextualize procurement prices, we have retrieved consumer price indices and inflation data from the Slovenian open data portal (OPSI), spanning the last decade. This enables us to assess whether price changes in bids are consistent with broader market trends, or whether anomalies may suggest irregular pricing strategies.

Table 6. Data obtained by OPSI portal by SNEP for Pilot 2.

Number of datasets identified	1
Partner	SNEP
Dataset description	Consumer price indices and inflation data from the Slovenian open data portal (OPSI), spanning the last decade.
Related WP/Task	WP5 / Task 5.2

Will you reuse any existing data? If YES, how?	Yes. Analysis of procurement practices and business relationships, identification of patterns and potential anomalies.
Methodologies for data collection / generation	Gathered via API, and FTP, CSV2SQL transformer.
Data type(s)	Text
Data format(s)	CSV
Data storage	https://podatki.gov.si/dataset/surs0400600s https://podatki.gov.si/dataset/surs0400605s For the purposes of CEDAR, data is downloaded and stored at SBC infrastructure.
Expected size of the data	100 MB
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	End users for monitoring procurement process. Data Scientist interested in evaluating algorithms and methodologies aimed at identifying patterns and anomalies.
Data access	Public data.
Data privacy	Public data.

2.1.3 Pilot 3: Transparent Management of Foreign Aid for Rebuilding Ukraine

The datasets related with Pilot 3 is led by YC and the activities are performed in the scope of Task 5.3. Since the initial submission of the Data Management Plan (D7.2 delivered at M06), the Ukrainian pilot led by YC has continued to rely on the datasets initially described. During the reporting period, the focus was concentrated on data enrichment, schema evolution, and testing in real-world scenarios.

YC has identified 14 datasets that are key for the Pilot 3 and are essential for effective monitoring of public procurement, compliance, and risk analysis. These include comprehensive information from Ukrainian state registries, tax data, court decisions, and sanctions lists. The datasets also cover entities with outstanding tax liabilities and records of violations in economic competition which are relevant for several groups, including public authorities, regulatory authorities, and data scientists.

In addition, Pilot 3 incorporates analytical indicators developed by YC— including express analysis, financial scoring, market scoring, and credit scoring — to further support reaching Pilot's goals in preventing fraud, supporting transparency and informed decision-making in public procurement.

In addition to these datasets, this pilot also actively integrates analytical partner outputs. ART contributes datasets derived from social media analysis and machine learning. This data is focused on Ukraine-related public signals of risk and corruption and flagging suspicious behaviour or relationships between tender participants and PEPs.

All of the data identified by partners is hosted in partners' secure environments, stored in structured formats and is accessible via API through an API key for authorisation, allowing for real-time integration. Schema documentation is maintained for all dynamic fields.

Since the initial Data Management Plan submission, two new complementary datasets have been added to Pilot 3 to enhance the granularity and real-time visibility of public procurement in Ukraine, particularly within the context of post-war recovery and donor funding.

DREAM Procurement Data: contains structured information on reconstruction-related procurement projects across Ukraine. The dataset includes details on project names, types of works or services, contracting authorities and suppliers, implementation stages, geographic location, etc. This dataset is used to track and assess the transparency and progress of post-war recovery efforts. It supports risk analysis, scoring, and the development of Pilot 3 for monitoring the allocation and implementation of foreign aid. The dataset is particularly valuable for stakeholders such as donor institutions, public authorities, civil society watchdogs, and journalists involved in oversight of Ukraine's reconstruction.

Table 7. DREAM data provided by YC for Pilot 3.

Number of datasets identified	1
Partner	YC
Dataset description	Procurement data on reconstruction-related projects in Ukraine, including project scope, location, type, and implementation status.
Related WP/Task	WP5 / Task 5.3
Will you reuse any existing data? If YES, how?	The Dataset is currently used by all public Ukrainian authorities for transparency
Methodologies for data collection / generation	Gathered via API
Data type(s)	Different data types (numerical, categorical, textual)
Data format(s)	API
Data storage	https://docs.google.com/document/d/1ncXkBgLt5lT7nUUWPIYOaMwadDgQ_hcNYKi3nbXj1cM/edit?tab=t.0#heading=h.hs6dr9ypa5n1
Expected size of the data	API requests
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	All Ukrainian public authorities, donor organisations, oversight bodies, civil society, data analysts, other end users for monitoring rebuilding procurement processes.
Data access	Opendata
Data privacy	Opendata

Prozorro Procurement Data: provides comprehensive information about all public procurement procedures conducted across Ukraine. The data includes unique tender identifiers, information about buyers and suppliers, contract values, procurement types and procedures, awards, cancellations, and associated documentation. This dataset is crucial for building a general overview of procurement practices, detecting anomalies, monitoring the integrity of contracting parties, and comparing tender dynamics over time. It is of particular relevance to regulatory authorities, anti-corruption experts, policy analysts, and researchers interested in public sector accountability.

Table 8. Prozorro data provided by YC for Pilot 3.

Number of datasets identified	1
Partner	YC
Dataset description	Structured data on all public procurement procedures in Ukraine, including tenders, contracts, participants, values, and related documents.
Related WP/Task	WP5 / Task 5.3
Will you reuse any existing data? If YES, how?	The Dataset is currently used by all public Ukrainian authorities for transparency
Methodologies for data collection / generation	Gathered via API
Data type(s)	Different data types (numerical, categorical, textual)

Data format(s)	API
Data storage	https://prozorro-api-docs.readthedocs.io/uk/latest/index.html
Expected size of the data	API requests
Metadata and standards	No particular metadata or standards have been followed while collecting the data.
For whom might the dataset be useful?	All Ukrainian public authorities, anti-corruption analysts and data scientists, researchers, journalists, other end users for monitoring procurement processes.
Data access	Opendata
Data privacy	Opendata

2.2 Updated Types and Formats of Artefacts Generated / Collected

2.2.1 Synthetic Data Generation

In order to facilitate the Slovenian use case, a synthetic data generation method was developed utilizing Python, amalgamating rule-based logic with generative AI to facilitate technical validation and risk detection without the dependence on actual or sensitive procurement data. The dataset comprises structured .csv files that represent tenders, bids, and orders, incorporating 600 standard and 100 potentially fraudulent cases. Fundamental fields such as names, addresses, and dates were generated utilizing the Faker library, whereas more intricate, context-specific fields, such as service descriptions, technical specifications, and price estimates, were generated using the Gemma 2 27b large language model via prompt engineering, operated on an Ollama server equipped with dual NVIDIA RTX 4070 GPUs. Fraudulent scenarios were constructed based on four pivotal indicators derived from the Slovenian pilot and are distinctly labelled to facilitate targeted validation. The resultant dataset is entirely synthetic and anonymized, thereby constituting a valuable resource for researchers, developers, and policymakers engaged in procurement transparency and fraud detection. Comprehensive details regarding the data generation process, including schema definitions and prompt structures, are available in Deliverable 2.2.

2.2.2 Other Open Data Sources

During this period, we have also collected data from open data sources, including the DIAVGEIA site of Greek Government’s public spending services. We have collected and persisted in a relational database and an object storage, the following data records aligned with CEDAR activities:

- **Unique Announcement Number:** The unique identifier assigned to each published governmental act or decision by the open data portal (in our case, the Greek one, DIAVGEIA).
- **Government Institution:** The specific governmental body (e.g., Ministry of Finance, Ministry of Health, Independent Authority for Public Revenue, etc.) responsible for issuing the act or decision.
- **Act/Decision Title:** A concise and descriptive title of the uploaded document.
- **Subject Matter/Keywords:** Relevant keywords or categories that describe the content and purpose of the act or decision.
- **Date of Issuance:** The official date when the act or decision was formally adopted by the issuing institution.
- **Date of Upload:** The exact date and time when the document was uploaded to the "Transparency Portal."
- **Document Format and Digital Content:** The file format and the actual file (e.g., JSON, PDF, DOCX, TXT) in which the act or decision is published.

- **Digital Signature Status:** Confirmation of the presence and validity of the digital signature associated with the document.

The process of data collection has been performed through the development of scheduled tasks designed to automatically extract information from the DIAVGEIA Portal. The acquired data are systematically stored within a relational database as well as an object storage system. The indexing of this information facilitates the ability to search through the documents and precisely discern specific governmental information and documents based on multiple criteria (e.g., time, type of ministry, keywords, etc.), irrespective of the technical expertise of the end user.

This dataset allows for the evaluation of compliance by government institutions with national and EU legislation[1] and the online publication mandate. It also permits the extension of research activities beyond the CEDAR pilot tasks and enables the examination of other public datasets, aligned with detecting patterns or anomalies that may suggest delays in publication, inconsistencies in data management, or potential issues of concern related to governmental activities, national security, or other matters. Finally, by analysing the relationship between online publication and the indicators of transparency, corruption, and citizen engagement, it becomes possible to identify areas for improvement in the publication process, contributing to enhanced governmental transparency.

2.2.3 Artefacts and Access Rights

Since the initial Data Management Plan submission (e.g., D7.2), ART's data contributions for the Ukrainian pilot have evolved, focusing on schema enhancements, data enrichment, and integration into tools and technical components of WP4 and WP5. Details on the data operations (DataOps) applied are elaborated in Table 9.

Table 9. Artefacts and Supplementary Datasets

ART SM Artifact	The internal processes at ART for transforming the dataset, which previously included raw data from social media and raw search results, into structured JSON format have been solidified to ensure consistency. The most valuable entities were extracted, potential relations between them were described, and they were actively tested under tasks 4.1, 4.2, and 4.4. The dataset is internally usable despite its noisy nature and the high complexity of working with it. It contributed to the enrichment of the Art Scoring Artifact during internal ML algorithms development and data processing.
ART Scoring Artifact	This dynamic dataset, generated via continuous social media analysis using ART's proprietary ML algorithms (building under tasks 4.1, 4.2, and 4.4), identifies Ukraine-related risk indicators, suspicious activities, and relationships concerning tender participants. A key update to this artifact is that the data now includes company groups, which allow for working with official and unofficial relations between companies and their related people. Also, numerical scores of company activity on social media were revised to avoid aggregation, so as not to lose valuable information from particular social media pages. After applying various algorithms to extract data insights, the artifact offers a more interconnected and detailed view for advanced risk analysis. Data access to the artifact is provided through the dedicated API and is integrating into Ukrainian Pilot 3 (WP5, Task 5.3). On the other hand, the key artefacts generated or used in the context of developing and deploying CEA's ML-based anomaly detection system (SIGMO-IDS), and the corresponding access policies are exposed hereafter. Artefacts span public datasets, proprietary system components, and intermediary outputs created for deployment, testing, and evaluation.
Public Intrusion Detection Datasets	Publicly available datasets such as CIC-IDS-2017 and CIC-IDS-2018, used for training and validating ML models. These datasets are subject to open access under the terms defined

[1] http://elib.aade.gr/elib/view?d=/gr/act/2013/4210

	by the data providers (e.g., Canadian Institute for Cybersecurity). Used with proper attribution and data use compliance.
DataOps Pipeline Components	This comprises a set of custom software and scripts developed by CEA for traffic collection, pre-processing, threshold calibration, and anomaly detection. This includes data ingestion scripts, feature extraction tools, and ML training modules. Data pipeline components are mainly proprietary components tied to platform-specific deployment. They will remain internal or be released under a controlled license.
Trained ML Models	Trained versions of the intrusion detection models based on CIC datasets or internal traffic data. Models trained on public datasets may be shared under a research license. Models trained on internal or sensitive data (e.g., CEDAR traffic) will not be publicly released due to security concerns and risk of reverse engineering.
SIGMO-IDS Deployment Artefacts	These artefacts are associated to the following infrastructure elements: • Network scanners for CI/CD environments • Kafka topics (network_raw, ids_alerts) • Inference back-end and REST API services • Authentication mechanisms and configuration files Manipulation/test/exploitation of these elements is restricted to project partners. Deployment artefacts contain sensitive operational logic and may expose system architecture details.
Unit Tests and Dashboard Backend	This includes testing infrastructure and visual components used to monitor model performance and inference outcomes. Security-sensitive or infrastructure-bound components remain internal.

3 Implementation of FAIR Principles

As we detailed in Deliverable 7.2, CEDAR is supporting research data reuse by following FAIR principles, to make data Findable, Accessible, Interoperable, and Reusable. The previous deliverable outlined our FAIR strategy, that has accomplished the following results:

3.1 Findable Data

The project has undertaken substantial effort over its first year to capture a vast collection of datasets from the three participating pilots, as reported in the previous section. These datasets are made available internally to the technical project partners through the project internal data exchange platform (MS SharePoint repository). These files are registered and stored according to the metadata and naming conventions detailed in Deliverable 7.2.

In addition to the raw data sources, project partners have defined a knowledge graph model that harmonizes these data sources into a common set of reasoning abstractions. The data transformed under this model is processed and maintained in its current version by ENG, who provides access to the rest of the consortium through an API.

3.2 Accessible Data

While at this stage of the project there are no new CEDAR datasets that are ready to be made accessible the project is building the CEDAR Minimum Viable Dataspace (MVD) that will provide secure access to the exploitable datasets that are assessed to have value outside the internal use cases. Next section provides details on the progress of this component.

Moreover, the project is opening to the public its main research accomplishments through the publication of its core results in top tier international research publications. At the time of writing this deliverable, a total of 1 paper have been published disseminating the initial project results to the scientific community. Open access mechanisms have been followed to ensure maximum reach of these results. Similarly, the project is also producing a series of whitepapers that highlight and describes a number of characteristics, practices, as well as lessons that the project has learned during this time. These whitepapers serve the purpose of providing general guidelines for those who will embark in a similar effort to that of CEDAR.

3.3 Interoperable Data

The data gathered for each of the data pilots is stored with a set of dictionary terms that provide unambiguous multilingual translations for the core terms behind each pilot. Building on top of them, the CEDAR Data Model bridges the concepts among the different datasets into a single interoperable abstraction that is compatible with all the project scenarios. While the current version of the CEDAR Data Model has been curated specifically for these use cases, the project is also developing automated tools for data alignment that will greatly enhance interoperability across project data for analysis, as well as data harmonization across the CEDAR MVD.

3.4 Reusable Data

Regarding data reusability the project has made significant advances in the generation of synthetic datasets from the initial raw datasets, as described in Section 2.2 of this deliverable. This approach can greatly extend the reuse of the available data sources.

4 Research Outputs

The CEDAR research outputs during the reporting period demonstrate significant advancements across data managed in the frame of Continuous Integration and Continuous Deployment (CI/CD) scripts, the CEDAR Minimum Viable Dataspace (MVD), the CEDAR Data Model, and research data results on cryptocurrency analytics, security and intrusion detection methods.

Regarding the CI/CD & integration environment, as part of the contributions to the CEDAR Project, a secure and robust CI/CD (Continuous Integration and Continuous Deployment) pipeline has been established, in order to support the development and integration of research outputs. This environment leverages industry-standard tools and practices, including GitHub for source code management, Jenkins for automation, Harbor for container image management, Portainer for container orchestration, and Keycloak for role-based access control (and authentication). The CI/CD infrastructure is deployed in a cloud environment protected via VPN access and firewall restrictions, which helps ensure that all data exchanged during the build, test, and deployment processes remains within a secure, isolated network. The formats handled in this environment primarily include source code, Docker images, deployment manifests (YAML/JSON), Jenkins files, and logs, all of which are managed in a traceable and reproducible manner. While not all these formats are directly aligned with the CEDAR Data Schema, the pipeline supports the development and integration of components that do utilize CEDAR-aligned formats for research outputs.

As for the Integration with CEDS, the implementation of the CEDAR Minimum Viable Dataspace (MVD) utilizes the sovity Data Space Connector, an enhanced distribution of the Eclipse Dataspace Connector (EDC), as well as sovity's DAPS (Dynamic Attribute Provisioning Service) for identity and access management, which builds on Keycloak. The data formats supported by the connector are consistent with those of EDC and IDSA specifications, ensuring semantic interoperability and technical compliance. Data exchanged within the MVD will conform to the CEDAR Data Model (and the associated schema definitions), thereby ensuring full alignment with the CEDAR Data Schema. This includes the use of standardized metadata representations, and policy-enforced data sharing. The connector infrastructure enables structured, compliant, and traceable research data exchange among project partners and (potentially) external stakeholders, forming a foundational layer for trust and interoperability within the CEDAR platform's ecosystem.

With respect to the CEDAR Data Model, the first iteration was defined as a knowledge graph, and documented through detailed diagrammatic and textual description. This approach supports easy updates on short notice. This first iteration was considered for the MVP, envisioning an iterative process where the pilots might require extra features almost on the fly to accommodate aspects that were left out in the initial consideration of the use cases. Moving forward, next iterations of the Data Model will be developed under the standard of JSON-LD. This type of file will allow for the storage of Data Model with consideration for the Linked Data format. In turn, this will help to ensure that the relations defined amongst the different entities of the Data Model are preserved and more easily converted into the necessary elements that constitute the CEDAR knowledge graph.

Moving forward with the idea of the CEDAR Data Model, the development of Autonomous Data Alignment will be able to produce ad-hoc data models on the fly and store them in the above-mentioned JSON-LD format. This allows to have any possible requirements aligned with the data that is currently in use, ensuring that no entities or relations are part of the file but only those that would be actually required. Furthermore, since this data models can be produced on the fly, their storage can be deemed sometimes unnecessary, as their creation can be done on the spot. Nonetheless, the most relevant ones, as well as the main CEDAR Data Model, encompassing all of the data that has been provided by the pilots should remain as permanent staples to provide support for further development of the knowledge graph.

The Kriptosare tool, developed by VICOM, uses information from cryptocurrency transactions and addresses as input. It offers several key functionalities: (i) extracting statistics and metadata from cryptocurrency blockchains, including OSINT data; (ii) facilitating the traceability of funds across different cryptocurrency entities (Follow-The-Money); and (iii) applying machine learning models and graph analysis techniques to classify the behaviour of blockchain entities. For the latter, Kriptosare constructs an address-transaction graph based on a given cryptocurrency address and performs behavioural classification on the resulting structure. The Kriptosare tool has been extended to support FIAT transaction data as well. Specifically, it can construct transaction graphs from FIAT data, trace the flow of funds between wallets, and extract indicators that may signal potential money laundering activities via anomaly detection techniques. This enhancement equips the CEDAR Project with capabilities to analyse and investigate transactions related to public procurements and specific companies.

As part of the development of CEA's ML-based intrusion detection system (SIGMO-IDS), substantial progress has been made on the end-to-end detection pipeline. This includes the full chain from network traffic collection to real-time inference and alert generation. The system supports modular pre-processing of traffic data, feature extraction tailored to anomaly detection tasks, and the integration of machine learning models capable of identifying suspicious patterns in near real-time. A key feature under finalization is the inclusion of a parameterizable threshold selection mechanism, allowing the detection logic to be calibrated according to the operational context or desired sensitivity levels. In parallel, a comprehensive suite of unit tests is being implemented to ensure the reliability and correctness of the core detection components. The backend of a monitoring dashboard is also under development, providing a foundation for user interaction, system status visualization, and integration of alert summaries. These components are further reinforced by an API layer and authentication mechanism, enabling secure and structured communication between the detection engine and external services. The proposed model supports the integration of lightweight network scanners within CI/CD environments, particularly targeting CEDAR platform components that expose REST APIs and are considered high-value assets. These scanners will collect and stream network traffic metadata to a Kafka topic (`network_raw`), ensuring efficient and decoupled data ingestion. The central IDS engine, hosted at CEA, will continuously process this streamed data to detect anomalies and publish alerts to a separate Kafka topic (`ids_alerts`), enabling downstream security automation.

5 Data Security

A detailed risk assessment focusing on data protection and information security was conducted under WP4. This analysis identified potential vulnerabilities, compliance risks, and mitigation strategies related to the handling of project data.

Based on the outcomes of the risk assessment activity in WP4, it is particularly worthwhile to point out that data security is a critical aspect of the project's data management strategy, requiring attention across all stages of the data lifecycle. The following measures are to be considered and proportionally adopted, depending on the nature and sensitivity of the data:

- **Access Control:** Role-based access mechanisms have been applied to restrict data access to authorized personnel only. Logging and audit trails may be established to monitor and review data access and usage.
- **Data Storage:** Data are being stored on secure, encrypted platforms, whether using institutional infrastructure or approved cloud services, ensuring alignment with standards such as ISO/IEC 27001 and GDPR.
- **Data Transmission:** All data transfers will be conducted via encrypted protocols (e.g., HTTPS, SFTP) to maintain integrity and confidentiality during transit.
- **Backups and Recovery:** Regular and automated backup procedures will be defined, with consideration for off-site storage and recovery mechanisms to ensure data availability in case of system failure or loss.
- **Sensitive Data Handling:** For datasets containing personal, sensitive, or confidential information, anonymization or pseudonymization techniques will be applied. Such data may also be subject to additional access restrictions and encryption.
- **Compliance and Training:** All team members involved in data handling will be made aware of applicable legal, regulatory, and institutional requirements. Training or reference materials will be made available to promote responsible data management.

These considerations, guided by the WP4 risk assessment, aim to ensure that security remains embedded in all aspects of data management, i.e., from collection and storage to sharing and long-term preservation.

6 Ethics

6.1 Ethics Requirements in Research

Ethics considerations can entail a great variety of topics. In CEDAR's DMP, research activities are the focus of ethical requirements explored. Since the proposal stage, relevant publications were evaluated, especially those focusing on best practices in the European funding scheme. From this, it was possible to build a list of ethics requirements to be observed by all partners in the CEDAR consortium, understanding the different goals of data management and ethics requirements, namely the internal data management for the project and the research activities developed in the pilots and technical tasks. While in this document, a summary of these considerations is presented, this Deliverable should be read next to the other relevant deliverables in the project, especially D7.3 (M06), and the future D7.5 and D7.7.

6.1.1 Regulation (EU) 2021/695 establishing Horizon Europe

Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination (Horizon Europe Regulation) is of utmost importance for ethics considerations in CEDAR. It establishes rules for the "participation and dissemination concerning indirect actions" in the Horizon Europe projects.[2]

Article 19 of the Regulation is solely dedicated to Ethics and highlights that all actions funded by Horizon Europe must comply with ethical principles and relevant law, which includes the European Charter of Fundamental Rights (Charter) and the European Convention on Human Rights (ECHR) and its Supplementary Protocols. Different points of attention are mentioned in Article 19(1), namely the principle of proportionality, the need to ensure protection of the environment and high levels of human health protection, and the rights to privacy, protection of personal data, physical and mental integrity, and non-discrimination.

Nonetheless, the focus of the Horizon Europe Regulation is actions to be taken by participants even before the beginning of a funded project (*ex-ante* protection), as means to guarantee an ethics-by-design approach. CEDAR has complied with all requirements, with special attention to the ethics self-assessment completed in the proposal stage. During this period, it was already mapped and highlighted that activities during the project would not involve human participants or the use of human embryos or embryonic stem cells. The biggest points of attention mapped during the initial ethics self-assessment were the use of Artificial Intelligence (AI) and the participation of non-EU countries, including for activities involving personal data processing activities.

Stemming from Article 19(2) Horizon Europe Regulation, CEDAR still had to receive a confirmation that the activities carried out outside the Union (the Ukrainian pilot) would have been allowed in a Member State. Previous to the research activities, an initial assessment was already performed guaranteeing that the activities to be conducted outside the EU would comply with all the ethical requirements for research as in the EU. Beyond that, throughout the project development, the participation of third country partners and an external pilot have been points of attention and specific considerations are brought to ensure the best approach following European principles, values and freedoms, including the proportionality balance with non-discrimination. As established in the "Ethics Summary Report", the project includes personal data processing by Ukraine, a non-EU country, what justified the recommendation for an ethics check to guarantee "that the necessary safeguards and steps were taken to confirm that EU laws and regulations are applied, information is not misused or exploited, and the possible vulnerability of certain demographics is not taken advantage of".[3]

And, as stated in the same document, this verification step should take place "after the initial catalogue of data and before the public and private data collection", during T2.1, and no repetition was needed. This was performed during the initial months of T2.1, and the results can be found in D2.1 (Initial Data Catalogue and Data Preparation Methods).

[2] Article 1(1) Horizon Europe Regulation.

[3] CEDAR's Ethics Summary Report.

6.1.2 European Code of Conduct for Research Integrity

The ALLEA European Code of Conduct for Research Integrity[4] is an essential piece for research activities developed under the Horizon Europe scheme, reinforcing the importance of trustworthiness in research results and considers important developments in law and best practices in the academic sector. The document lists non-exhaustive fundamental principles of research integrity, including:

- Reliability;
- Honesty;
- Respect; and
- Accountability.

For observing the full potential of these principles, the Code also describes good research practices. Especially in Data Management, the Code highlights that it is essential that research organisations:

- Ensure appropriate stewardship, curation, and preservation of all data, metadata, protocols, code, software, and other research materials for a reasonable and clearly stated period;
- Ensure access to data as open as possible and as closed as necessary, and, where appropriate, in line with the FAIR principles;
- Are transparent about how to access and gain permission to use data and other research materials;
- Acknowledge data as legitimate and citable products of research;
- Ensure that any contracts or agreements relating to research results include equitable and fair provisions for the management of their use, ownership, and protection under intellectual property rights;
- Inform research participants about how their data will be used, reused, accessed, stored, and deleted, in compliance with the GDPR[5].

The Code is also an important source of understanding misconducts and other unacceptable practices, what can be illustrated by “fabrication, falsification, or plagiarism (the so-called FFP categorisation) in proposing, performing or reviewing research, or in reporting research results”.[6] Nevertheless, as already highlighted in the Code of Conduct, misconducts against research integrity are non-exhaustive and can be presented in different forms. Transparency about the expected behaviours is, then, of utmost importance. Also, in case of any suspicious action, this is to be assessed in a fair way by the consortium and possible external advisors to ensure a case-by-case evaluation considering the best practices and the specific circumstances applied to the scenario.

6.2 Personal Data Protection

Personal data processing is essential to CEDAR’s goals, including both internal activities of management of partners and participants and personal data use for the research goals.

As detailed in D7.3, any data processing activity involving personal data should comply with the applicable requirements set by law. By understanding that any personal data processing entails risks to data subjects, CEDAR aims to adopt systems and procedures to mitigate those risks. For this, the General Data Protection Regulation (GDPR) is used as a guidance for best practices and legal requirements. The Regulation is based on three significant elements for personal data protection: controllers' accountability, data subjects’ rights, and supervision by independent authority [2].

[4] Allea, European Code of Conduct for Research Integrity. Revised Edition (2023) http://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf

[5] Until the moment, CEDAR does not foresee any human participation in the project.

[6] Allea, European Code of Conduct for Research Integrity. Revised Edition (2023) http://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf

These elements are essential to guarantee the individual control over their personal data and to ensure oversight in case of infringement of data protection rules. Additionally, as a translation of the system established by the GDPR, it is worth highlighting the data protection principles, detailed in D7.3, and considered in the DMP's elaboration and on guidelines developed under WP7.

6.3 Non-personal Data Governance

Beyond personal data, CEDAR Project also considers the best practices related to non-personal data processing. As already highlighted in this Deliverable, these include the FAIR principles, open data considerations, and large exploration of research results. Beyond those, as mentioned in the ethics self-assessment, CEDAR will also consider the principles, values and directions established by the European Data Strategy [4][3] and its composing norms, especially the Data Governance Act [4]. Other pieces of legislation complete the strategy, and the data governance rules set by the EU, especially the Open Data Directive [5]. A detailed study and guidance based on these norms was presented in D7.3 to guarantee a common approach of all partners in CEDAR in the non-personal data governance.

6.4 Artificial Intelligence Use

Artificial intelligence (AI) use can raise different ethics considerations and risks to ethical standards. This assessment depends on the risks to fundamental rights and freedoms to individuals and even impacts to common goods and shared interests. As CEDAR intends to work with AI systems, the ethics partners are following up established[7] and new instruments[8] to understand the best use of AI in different stages, from developing a system to applying and offering an AI solution. In either way, it is of utmost importance for CEDAR to guarantee the transparency of AI use in the project. Transparency should be guaranteed at different levels and to different stakeholders impacted by CEDAR's outputs, including participants of pilots, future users of the project's results, and society as a whole. Clarity around AI use allows better oversight of the results and uses, which can be illustrated by facilitated audits.

[7] E.g., European Commission, Ethics Guidelines for Trustworthy AI from the Independent High-Level Expert Group on Artificial Intelligence. Brussels, 8.04.2019. Retrieved from: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

[8] E.g., Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

7 Conclusions

This deliverable documents the activities undertaken and the methodology employed by the CEDAR Consortium to achieve efficient and transparent data management within the project up to M18. Our efforts have concentrated on maintaining the available, collected, utilized, and synthetically generated datasets in a current state to ensure transparency, reproducibility, and collaboration within the CEDAR Project and the broader scientific community.

Through the establishment of the DMP Register, we have ensured that data is systematically enumerated, organized, preserved, and complies with the FAIR principles, applicable both to the current project and future initiatives. In the forthcoming period, we intend to maintain the currency of the DMP Register and enrich it with additional datasets as the CEDAR Project progresses.

8 References

[1] Data Management Plan in Horizon Europe. Available at: https://enspire.science/data-management-plan-in-horizon-europe/

[2] Felix Bieker. 2022. 34 The Right to Data Protection. Individual and Structural Dimensions of Data Protection in EU Law. The Hague, The Netherlands: Springer. Gloria González Fuster. 2014. 16 The Emergence of Personal Data Protection as a Fundamental Right of the EU. Springer Science & Business.

[3] European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data. Brussels, 19.2.2020. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.

[4] 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). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868.

[5] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng

Funded by the
European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101135577