



CENTREPEACE

DATA MANAGEMENT PLAN UPDATE I

Deliverable 10.2

Adriana Ilavská, Marián Koval



**Funded by
the European Union**

OUTPUT SUMMARY Project Information	
Project Title	Central and Eastern European Security Cooperation Cluster
Project Acronym	CENTREPEACE
Call Identifier	HORIZON-WIDERA-2023-ACCESS-02-01
Contract Number	101159051
Starting Date	01.06.2024
End Date	31.05.2027
Web-Site Address	www.centrepeace.cz
Coordinator	Masaryk University
Management Team	Zinaida BECHNÁ, Martin CHOVANČÍK, Marián KOVAL
E-mail	centrepeace@fss.muni.cz
Phone Number	+420 459 494 573

Deliverable Title	Data Management Plan Update I
Deliverable Number	D10.2
DMP Version	1.1
Work Package	WP10 Project Management: the Engine of collaboration
WP Leader	MUNI
Nature	R
Dissemination	PU
Editor (s)	Adriana ILAVSKÁ, Marián KOVAL
E-mail (s)	ilavska@fss.muni.cz , koval@fss.muni.cz
Phone Number	+420 549 494 573
Date of Delivery	30.10.2025

Verification by	Anne DUQUENNE, Maija HARJU
Date	24.10.2025
Validation by	Zinaida BECHNÁ
Date	30.10.2025

Disclaimer: Views and opinions expressed are 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.

This Data Management Plan was created in Data Stewardship Wizard



Contents

Version changes	4
1. Data Summary	5
2. FAIR Data	10
2.1. Making data findable, including provisions for metadata	10
2.2. Making data accessible	11
2.3. Making data interoperable	14
2.4. Increase data re-use	15
3. Other research outputs	18
4. Allocation of resources	19
5. Data security	20
6. Ethics	22
7. Other issues	23

Version changes

Version 1.0	Initial version
Version 1.1	Chapter 1. Data Summary • Added SIPRI Arms Transfers Database • Added V-DEM (Varieties of Democracy) Dataset • Media Article Datasets got updated • The part “to whom might data be useful” got updated Chapter 2. FAIR Data • Data Collection Documentation got updated • The subchapter Metadata – Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software” been updated Chapter 4. Allocation of resources • The part about preservation ensurement got updated

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

We will re-use existing datasets to supplement our primary research data and enhance our analysis. The datasets considered for re-use, in this early stage of the research, are:

- **Eurostat database**

It is available via: <https://ec.europa.eu/eurostat/web/main/data/database>.

Owner of this dataset: European Commission, Eurostat Joseph Bech Building 5, rue Alphonse Weicker L-2721 Luxembourg.

The dataset can be used in the provided format without any conversion needed.

We will keep a copy of the dataset and make it available with our results for reproducibility.

- **World Bank Open Data**

It is available via: <https://data.worldbank.org/>.

Owner of this dataset: THE WORLD BANK 1818 H Street, NW Washington, DC 20433 USA Tel : (202) 473-1000.

The dataset can be used in the provided format without any conversion needed.

We will keep a copy of the dataset and make it available with our results for reproducibility.

- **EU arms export data**

It is available via: <https://github.com/caatdata/eu-arms-export-data>. It is used in the project.

Owner of this dataset: Campaign Against Arms Trade (CAAT) Tel: 020 7281 0297 Email: enquiries@caat.org.uk.

The dataset can be used in the provided format without any conversion needed.

We will download or get a copy, as the dataset may get updated in the future; therefore, we need to make a snapshot.

Only a part of the dataset will be used; any filtering or selection will be well-documented.

- **SIPRI Arms Transfers Database**

It is available via: <https://armstransfers.sipri.org/ArmsTransfer/ImportExport>.

Owner of this dataset: Stockholm International Peace Research Institute (SIPRI), Signalistgatan 9, SE-169 72 Solna, Sweden.

The dataset can be used in the provided format without any conversion needed.

We will download and keep a snapshot of the dataset to ensure reproducibility, as it may be updated over time.

Only a part of the dataset will be used; any filtering or selection will be well-documented.

- **V-Dem (Varieties of Democracy) Dataset**

It is available via: <https://www.v-dem.net>. It is used in the project.

Owner of this dataset: V-Dem Institute, Department of Political Science, University of Gothenburg, Sprängkullsgatan 21, 405 30 Gothenburg, Sweden.

The dataset provides a comprehensive set of democracy indicators across countries and time, covering political institutions, participation, and civil liberties.

The dataset can be used in the provided format without any conversion needed. We will download and store a dataset version to ensure reproducibility and transparency. Only relevant variables will be selected for our analysis, and the selection process will be fully documented.

We consider using the datasets to provide essential context for understanding regional cooperation and the socio-economic and political factors influencing the cooperation before gathering our own data. These datasets will help us identify key patterns, trends, and relationships that shape regional dynamics and provide a broader framework for our analysis. By leveraging this information, we aim to ensure that our research is grounded in reliable, well-established data as we move forward with further data collection.

What types and formats of data will the project generate or re-use?

At this very early stage of research, we plan to re-use the existing datasets described above while also generating two original datasets of our own; these will be described to the extent possible given the current phase of the project.

Generated Data:

- **Expert Interviews:** Dataset of document-based data consisting of transcribed, anonymised interviews and coded results derived from a series of interviews conducted with experts in the relevant fields in TXT and DOCX formats.

- **Media Articles Dataset:** Dataset of document-based data collected articles in both their original languages and versions machine-translated to English, in TXT format, with metadata of articles (that are originally included in text files) extracted to separate files in CSV, format. The dataset will comprise media articles collected from Factiva, a database that selectively curates high-quality national sources, covering the period from 2004 to 2024 and spanning multiple European countries. To ensure consistency and reliability, the collection process involves documented search queries tailored to capture relevant media content, which will be included in the metadata file.

Re-used Data:

- **Eurostat database:** Dataset of tabular quantitative data consisting of economic, demographic, and social statistics in CSV format.
- **World Bank Open Data:** Dataset of tabular quantitative data consisting of global statistics, in CSV formats.
- **EU Arms Export Data:** Dataset of tabular quantitative data documenting arms exports, trade partners, and military rankings in CSV and XLSX format.
- **V-dem Dataset:** Dataset of tabular categorical data containing global indicators of democracy, political institutions, civil liberties, and governance quality, in RData or CSV format.
- **SIPRI Arms Transfers Database:** Dataset quantitative data documenting international transfers of major conventional weapons, including detailed information on arms imports and exports, suppliers, recipients, and categories of military equipment, in CSV format.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

- **Expert Interviews:** To gather qualitative insights from experts, supporting our research into regional perspectives on the defense industry.
- **Media Articles Dataset:** To analyse media coverage from 2004 to 2024 across multiple European countries, providing context and trends relevant to the defense sector.
- **Re-used Datasets:** To incorporate reliable statistical and political data from multiple authoritative sources that enhances our analysis and supports our findings.

What is the expected size of the data that you intend to generate or re-use?

The expected data size is moderate:

- **Expert Interviews:** Up to 5 GB, primarily due to text-based transcripts.
- **Media Articles Dataset:** The final dataset contains plain text files, resulting in a moderate size of up to 500 MB.
- **Re-used Datasets:** As the data are readily accessible from public curated databases, even in cases of high volumes, the size will remain manageable, with an estimate of up to 10 GB.

What is the origin/provenance of the data, either generated or re-used?

Generated Data:

- **Expert Interviews:** Conducted by our research team with experts in relevant fields.
- **Media Articles Dataset:** Collected from Factiva, a curated database of high-quality national media sources.

Re-used Data:

- **Eurostat, World Bank:** Official statistical data from reputable international organizations.
- **EU Arms Export Data:** Data are from Official Journal of the European Union annual reports on the European Union Code of Conduct on Arms Exports.
- **SIPRI Arms Transfers Database:** Data are from the Stockholm International Peace Research Institute, which compiles official information on international arms transfers from national reports, open-source materials, and expert assessments.
- **V-Dem:** Data are collected and maintained by the V-Dem Institute and are based on a global expert survey and rigorous methodological standards to measure democracy and political institutions.

To whom might your data be useful ('data utility'), outside your project?

Academics and Researchers: Scholars in defence studies, security studies, and Central and Eastern European (CEE) regional research can use the expert interview dataset for qualitative insights into security perspectives and strategies within the region. The media articles

database provides a rich source for analysing discourse on the defence industry and security policies across European countries, particularly in the context of CEE's role in EU security following Russian aggression in Ukraine. While the full articles cannot be shared due to Factiva's licensing policy, we will provide detailed documentation, search strategies, and processing steps to enable others to reconstruct and utilize the dataset for their own research purposes.

Policy Makers and Government Agencies: The interview dataset offers practical insights from experts on regional security strategies, while the media database highlights trends and public discourse on defence and security. These datasets can inform evidence-based decision-making and policy development, especially regarding the integration of CEE perspectives into EU security frameworks.

Non-Governmental Organizations and Think Tanks: The datasets enable organizations to analyse arms trade narratives, regional security priorities, and public sentiment on defence industry issues. The media database is particularly useful for exploring how CEE countries are portrayed in security debates, while the interviews offer direct input from experts on key challenges and opportunities for regional security cooperation.

2. FAIR Data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

The generated data will use the persistent identifier, either Digital Object Identifiers (DOIs) or stable URLs assigned through the repository of Masaryk University.

Stable URLs are sufficient for datasets that are intermittent and which are not the main outputs of the project. These datasets serve supplementary purposes and do not require the same level of formal identification as primary outputs. Key datasets will be assigned DOIs to ensure their discoverability, persistence, and citation across platforms.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed?

In our field of research, there is no widely accepted or standardized metadata framework. However, we have a clear understanding of the metadata needed to enable others to work with our data, and we will adopt a tailored approach to ensure the data is well-documented and reusable.

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

Given the absence of standardized protocols or lab notebooks in our field of research commonly found in other scientific disciplines, we will adopt a comprehensive and structured approach to document the origin and all processing steps of our data.

- **Data Collection Documentation:** Each dataset will include detailed metadata specifying its source—whether primary data from interviews or secondary data from the media database. We will document the methods and tools used for data collection, including the date and other relevant information, to the extent anonymization of interviews will allow. For the Media Articles Dataset, we will document the search queries used for article inclusion, ensuring that the dataset can be transparently reconstructed and replicated.
- **Data Processing and Transformation Logs:** Detailed logs will be maintained for all data processing steps, such as cleaning, filtering, coding, and aggregation. We will document the software and tools used, along with their versions and configurations.
- **Data Analysis Workflows:** We will provide clear descriptions of the analytical methods and statistical techniques applied, including the rationale and assumptions. Intermediate outputs will be documented to ensure each step can be traced and verified.

- **Regular Audits and Reviews:** Periodic reviews will be conducted to ensure the completeness and accuracy of provenance documentation. Peer reviews will be involved to provide additional validation and ensure transparency.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Relevant keywords will be included in metadata to enhance discoverability and facilitate potential reuse by other researchers.

Will metadata be offered in such a way that it can be harvested and indexed?

In the current stage of the research, we do not plan to structure metadata in a form compatible with any particular harvesting protocol, but they will be openly available including instructions on how to get access to the data.

2.2. Making data accessible

Repository

Will the data be deposited in a trusted repository?

The data will be stored in the trusted institutional repository of Masaryk University, which is suitable for the type of data we are working with and meets all requirements for data accessibility, security, and long-term preservation. This repository is a secure platform designed to handle research data and is readily available for the project's needs. In addition, we will manage access rights to the data based on involvement in the project and the nature of the data. During the project, data will also be securely stored on SharePoint/OneDrive, which is serviced by MUNI experts and equipped with all necessary protections.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

No further arrangements are necessary at this stage; however, should we work with sensitive data in the future, we will make additional arrangements to ensure appropriate access controls, which are fully supported by the institutional repository.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

The institutional repository ensures that data is assigned a stable URL, which functions as a persistent identifier for accessing the data. This stable URL resolves directly to the corresponding digital object within the repository, ensuring reliable access and referencing over time.

In case of further publications of the created datasets, assigning DOIs will be ensured to enhance their discoverability and citation.

Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions.

Not all data can be made openly available due to legal and ethical considerations:

- **Expert Interviews:** Transcripts cannot be shared openly to protect participant confidentiality. Anonymised versions will be shared under restricted access conditions, ensuring that confidentiality is preserved. The results of the analysis derived from these interviews will be shared openly.
- **Media Articles Dataset:** Full-text articles cannot be openly shared due to copyright restrictions. However, metadata describing the articles and the results of the analysis will be made available in compliance with Factiva's licencing terms to ensure transparency and support further research.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No embargo is planned. Data will be shared as soon as possible, respecting all legal and ethical constraints.

Will the data be accessible through a free and standardized access protocol?

Data will be accessible through standardized protocol HTTPS, which is provided by the institutional repository. Given the manageable volume of data in the project, no specialized protocols for large-scale data transfers are required and HTTPS ensures secure and reliable access to the data while maintaining user privacy and data integrity during transfer. The HTTPS protocol, supported by the repository, is sufficient to handle the anticipated access and download needs for users.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

We have a consortium agreement in place that governs the management of Intellectual Property among the project partners. This agreement ensures that data usage rights and restrictions are clearly defined and respected.

Data that can be published openly will be made available during the project through appropriate repositories, ensuring compliance with FAIR principles.

For versions of datasets which cannot be published openly, access will be provided through secure channels to authorized users. After the project ends, restricted data will remain accessible under the same conditions through the institutional repository, ensuring that access continues to align with legal, ethical, and contractual obligations.

How will the identity of the person accessing the data be ascertained?

Access to the data will be verified through SharePoint, where all users are required to have authenticated accounts. This ensures that only authorized individuals can access the data, maintaining security and accountability throughout the project.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

There is no need for a data access committee. All data that can be made open will be openly accessible. Data with restrictions will only be available under contractual agreements with project partners, ensuring compliance with legal and ethical requirements.

Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

Metadata will be openly available and licenced under a public domain dedication CC0.

Will metadata contain information to enable the user to access the data?

The metadata will contain instructions on how to get access to the data.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data will remain available and findable for a minimum of 5 years after the end of the project. The metadata will be guaranteed to remain available after data is no longer available.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The documentation will include detailed information about the software used to access or process the data. For licensed software that is not open source, users will need to ensure access to the software independently. However, all data will be provided in formats that can be accessed using open-source software. Additionally, if any operations or analyses are performed using open-source software, the corresponding code will be provided. In cases

where specific parts of the code contain elements that cannot be shared due to data provider policies, these sections will be clearly redacted and accompanied by explanatory notes to ensure that the overall process remains understandable and reproducible.

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In these early stages of the project, we have not yet identified all the specific vocabularies, standards, and methodologies that will be used to ensure data interoperability. However, we are committed to making our data as interoperable as possible to facilitate data exchange and re-use both within and across disciplines.

To this end, we will use the **ISO 3166** standard for country codes to ensure consistency and interoperability when referring to geographic data. For other aspects of data standardization, we will assess and adopt appropriate vocabularies and standards as the project progresses, particularly those widely recognized in relevant fields. Wherever possible, we will align with community-endorsed interoperability best practices to ensure that our data can be effectively integrated with other datasets.

While the character of the data is specific and may not fit perfectly within existing frameworks, we will actively explore how our data can be aligned with established vocabularies or mapped to common standards.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

If we need to use uncommon or generate project-specific ontologies or vocabularies due to the unique nature of our data, we will strive to ensure their interoperability by providing mappings to more commonly used ontologies and standards wherever possible. Additionally, any project-specific ontologies or vocabularies we develop will be openly published to allow others to reuse, refine, or extend them.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

Our data will include qualified references to other datasets to enhance its usability and contextual understanding. These references will be documented in the metadata to ensure the traceability of connections to related data sources.

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

We will provide comprehensive documentation to validate data analysis, and this documentation will be included in the repositories alongside the data to ensure easy and seamless access for users. All documentation will be updated as needed. Updates will be included in the repository alongside the data, allowing users to access it.

- **Expert Interviews:** For the expert interview dataset, will include detailed methodological documentation outlining the interview protocols and anonymization processes. A codebook will accompany the data to explain the coding framework, variable definitions, and thematic categories used in the analysis.
- **Media Articles Dataset:** For the media articles dataset, a readme file will describe the search queries. Additionally, we will include metadata detailing the publication sources, dates, languages, and any categorization applied during processing.

All datasets will include data cleaning logs that document steps taken to prepare the data for analysis, ensuring transparency and replicability. For any operations performed using software, we will provide scripts or code and clearly annotate these to explain their functionality. Units of measurement and other relevant technical details will also be included to ensure users can accurately interpret and re-use the data.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Whenever possible, our data will be made freely available in the public domain to permit the widest possible re-use. Data that can be openly shared will be licensed under the Creative Commons Attribution-NoDerivatives (CC BY-ND) license, which ensures that the data can be reused as long as proper citation is provided and the data are not edited or altered.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

The data that we can make openly accessible will be fully usable by third parties, provided they cite the data appropriately. These datasets will be licensed under terms that encourage reuse while ensuring proper attribution. However, data that cannot be made public due to legal and ethical considerations will not be available to third parties. Such restricted data will only be accessible to project partners in accordance with the established agreements.

Will the provenance of the data be thoroughly documented using the appropriate standards?

We will implement a comprehensive and structured approach to data quality assurance, ensuring accuracy, consistency, and reliability at every stage of the data lifecycle.

Describe all relevant data quality assurance processes.

Given the absence of standardized protocols in our field, we will develop tailored documentation and processes to maintain quality assurance. Key measures include:

- **Data Collection Documentation:** Each dataset will be accompanied by detailed metadata specifying its source—whether primary data from interviews or secondary data from media articles, official reports, or databases. We will document the methods and tools used for data collection, including dates, locations, and contexts, to ensure transparency.
- **Data Entry and Cleaning:** Data will be systematically checked for accuracy, completeness, and consistency.
- **Processing and Transformation Logs:** Detailed logs will document processing steps, including cleaning, coding, and aggregation. The software and tools used, along with their versions and configurations, will also be recorded to ensure transparency.
- **Validation of Analytical Workflows:** For critical steps, duplicate workflows will be implemented to allow double-checking by more than one team member. This process will include peer review of steps to ensure accuracy, consistency, and the minimization of errors.
- **Data Analysis Workflows:** We will provide clear descriptions of the analytical methods and statistical techniques applied, including the rationale, assumptions, and intermediate outputs for verification.
- **Documentation Practices:** Structured metadata will be supplemented with narrative descriptions and contextual information for a comprehensive understanding of the data lifecycle. Annotations and comments within scripts and documents will clarify the purpose and context of each processing step.
- **Regular Peer Reviews:** Periodic reviews will be conducted to ensure the completeness and accuracy of the data and its documentation.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

Beyond the data itself, the research outputs from this project will include reports, articles, and policy papers, all of which will adhere to rigorous quality assurance processes. These outputs will undergo a standardized peer review process.

These outputs will be openly accessible where possible, in alignment with the FAIR principles, and will be shared through suitable platforms or repositories to maximize their impact and

reach. Furthermore, their preparation will carefully consider ethical aspects, ensuring that no confidential or sensitive data is disclosed.



3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

In addition to data, the project will generate supplementary research outputs, possibly including project-specific ontologies, vocabularies or workflows developed to accommodate the unique nature of our data.

While no software, or physical outputs are anticipated, any project-specific ontologies, vocabularies or workflows will be openly published with appropriate documentation to support their reuse, refinement, and extension by others. These outputs will be shared online and linked to the project through DOIs or other persistent identifiers to ensure their discoverability and accessibility.

To align with the FAIR principles, all supplementary outputs will be provided with comprehensive documentation that explains their purpose, context, and applicability. Where applicable, mappings to commonly used ontologies and standards will be provided to enhance their interoperability. Open licenses will be applied to all shared outputs, enabling reuse while ensuring proper attribution.

The management and planning of other research outputs is done separately and is included as appendix to this DMP. Still, we benefit from data stewardship guidance (e.g. FAIR principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible. Making our data FAIR is not a cost that can be separated from the rest of the project.

No additional costs are expected for storage, archiving, or security as institutional repositories and infrastructure already meet the necessary requirements without incurring extra fees beyond regular institutional costs. We do not require any hardware or software in addition to what is usually available in the institutions.

We have a reserved budget for time and effort required to prepare data for publication for each team member. The Work Package 10 “Project Management: the Engine of collaboration” is dedicated to overseeing data handling, management, and reporting, comprises 6% of the overall budget. At this time, no additional costs are anticipated for ensuring that outputs are FAIR-compliant.

Who will be responsible for data management in your project?

CENTREPEACE Project Manager Marián Koval is responsible for implementing the DMP, and ensuring it is reviewed and revised in 18th month of the project and at the end of the project.

To execute the DMP, no additional specialist expertise is required.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

We will be archiving data for long-term preservation after the project in repositories ensuring preservation, but also data curation. The costs for archiving data will be paid out of departmental budgets from one or more of the project participants. The minimum lifetime of the archive is 5 years. The archival period can be extended. The decision whether or not to extend the renewal be based on the predicted use of the archived data. The renewal decision will be discussed within the Consortium to ensure alignment with project priorities and resource allocation. Data formats of data in cold storage will be upgraded if they become obsolete. Archived data will not be migrated to other storage media over time.

5. Data security

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

Robust provisions are in place to ensure the security of the data (especially sensitive data we are not to publish openly) throughout the project. These include:

- **Secure Storage:** Sensitive data will not be stored on individual computers, external hard drives, USB sticks, or other external media. Instead, all data will be securely stored in institutional trusted data centres and in SharePoint, which carry sufficient certifications for secure storage.

Raw data containing personal identifiers will be encrypted and stored on a secure server managed by the institutional repository or on SharePoint, which complies with GDPR standards and has robust access controls.

- **Access Controls:** Access to data will be regulated based on a collaboration agreement that specifies who can access which data. Only authorized personnel from participating groups and institutes will have access, with permissions based on their role in the project.

Access to the raw data with personal identifiers will be restricted to authorized team members only, using multi-factor authentication where possible.

- **Data Transfer:** All data transfers will occur via secure HTTP (HTTPS) to ensure data integrity and prevent unauthorised access during transmission.
- **Data Recovery:** Automated backups are integrated into SharePoint, ensuring that data stored there is protected against loss. The risk of information loss in the project is considered acceptably low, as a significant portion of the data is derived from external sources and can be readily restored if necessary.
- **Risk Mitigation:** The risks of data loss, leakage, or vandalised data are assessed to be low due to the robust security measures in place. In the unlikely event of such occurrences, the potential impact on the project or organization is minimal because a significant portion of the data originates from external sources and can be readily restored. For our gathered primary data, such as interview transcripts, extra attention will be devoted to creating secure backups, ensuring their protection and recoverability.

All project members have been instructed on both general and specific risks to the project, including data security protocols and best practices.

All data management practices will comply with **GDPR standards** to ensure the protection of data and ethical handling of information.

Will the data be safely stored in trusted repositories for long term preservation and curation?

The data will be securely stored in the trusted institutional repository of Masaryk University, which ensures long-term preservation, accessibility, and curation. Data will be archived for at least 5 years.

6. Ethics

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

There are ethics and legal aspects we have to consider when proceeding with data sharing in this project:

- **Expert Interviews:** Although the interview data will be anonymised for sharing, the process of data collection involves knowing the identities of the participants. This raises ethical considerations around confidentiality and informed consent. Participants will be fully informed about how their data will be used, stored, and anonymised, and they will provide explicit consent for these purposes. Anonymization will be implemented rigorously to ensure that individuals cannot be identified in the shared data. However, due to the nature of the data collection process, access to raw, unanonymised transcripts will be restricted to project team members, and such data will not be shared externally.

Otherwise, we will not collect any data connected to a person, i.e. "personal data".

- **Media Articles Dataset:** Media articles collected from Factiva are subject to copyright restrictions, which limit the extent to which full-text articles can be shared openly. To comply with legal obligations, only metadata, such as publication dates, sources, and other relevant descriptors, will be shared. Analysis results derived from these articles will also be made publicly available. Users interested in accessing the full texts will need to obtain them directly through Factiva or other authorized sources.

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

Informed consent for data sharing and long-term preservation will be included in interviews. As we will gather no personal data beyond participants' names, the consent will clearly specify how this limited information will be used, anonymised, and shared.

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

We will be using the following policies and procedures for data management:

- **FAIR principles**

In the project, Guidelines on FAIR Data Management in Horizon 2020 by European Commission

(https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf), are followed.

The framework was also adapted and is recommended to follow on the institutional level at Masaryk University (<https://openscience.muni.cz/en/fair-a-open-data/fair-data-a-jak-na-ne/hlavni-principy-fair>).

We will adhere to the FAIR principles to comply with the requirements of the European Commission but also because, these principles have been adopted, followed, and promoted by Masaryk University, aligning with our institution's commitment to open science and data sharing. In social science generally and also this research particularly, it is of utmost importance to share findings widely (as open as possible), as it fosters collaboration and aligns perfectly with our project goals. Given the aim of our research, the nature of the data we will be working with, and the available technological tools, we believe there are no obstacles to following the FAIR principles.