CHESS

Cyber-security Excellence Hub in Estonia and South Moravia

D5.2
Data Management Plan Update

Project Name	Cyber-security Excellence Hub in Estonia and South Moravia
Project acronym	CHESS
Grant agreement no.	101087529
Call	HORIZON-WIDERA-2022-ACCESS-04
Type of action	HORIZON-CSA
Project starting date	1 January 2023
Project duration	48 months
Deliverable Number	D5.2
Deliverable name	Data Management Plan Update
Lead Beneficiary	Masaryk University
Type	R – Document, report
Dissemination Level	PU – Public
Work Package No	WP5
Initial DMP submitted	30 June 2023
DMP update submitted	21 August 2025
Version	2

Funded by the European Union

Funded by the European Union under Grant Agreement No. 101087529. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Editors

- Václav Matyáš (MUNI)
- Zuzana Vémolová (MUNI)

Contributors

- Katarína Galanská (MUNI)
- Jan Hajný (BUT)
- Liina Kamm (Cybernetica)
- Lukáš Malina (BUT)
- Raimundas Matulevičius (UTARTU)
- Petr Švenda (MUNI)
- Jan Willemson (Cybernetica)

Reviewers

- Michal Růžička (Data Security and Management, MUNI)
- Pavel Šmerk (Open Science Methodologist, MUNI)

History of Changes

Page / Section	Nature of change and reason
p. 4	CHESS Consortium: Estonian Information Security Authority (associated partner) is substituted with the Estonian Association of Information Technology and Telecommunications (new associated partner).
p. 7 / Section 1	Introduction: A new introductory text in Section *1 Introduction* has been added.
p. 8 / Section 1.2	Organisation of the CHESS project: Updated according to the development in CHESS WPs. Link to CHESS Zenodo community added.
p. 9 / Section 1.3	CHESS Challenge Areas: Research focus updated and expanded for all Challenge Areas. A list of "Current Research" points has been added. The previous version covered only pilot research.
From p. 12 onward (starting at Section 1.4)	Minor edits to grammar and verb tenses were made to reflect the current status (e.g., future → present).
p. 13 / Section 2.1	Re-Use of Existing Data: Public reports on pentesting added among the types of data we are re-using.
p. 14 / Section 2.4	Size of Data: Revised text from "we do not expect any significant size of the data to be generated" to "we did not experience and do not expect..." to reflect project experience.
p. 14 / Section 2.5	Origin of Data: Revised text from "Some of the data is publicly available "to "Most of the data is publicly available ".
p. 14 / Section 2.6	Data Utility Outside of the Project: Changed "might" to "will" to reflect increased confidence in data usefulness. Added sentence "Nearly all software(s) shall be reusable for future research and allow for repeatability and verification of our research".
p. 16 / Section 3.1.2	Metadata and Their Standards: Two kinds of metadata added: • Software code with accompanying data. • Metadata related to project events. Some of our events have separate webpages.
p. 16 / Section 3.2	Making data accessible: The link to the theses registry of the University of Tartu changed.
p. 17 / Section 3.2	Will all data be made openly available? Replaced general statement on data availability with a detailed explanation of data sharing limitations related to CoinJoin privacy experiments (CA4) and company collaborations under NDAs. Clarified conditions under which aggregated or anonymised results are published.
p. 20 / Section 7	Ethics: Expanded ethics statement to include details about an ongoing ethics approval request at the University of Tartu for research involving usability of penetration testing reports, use of AI-generated content, and personal data collection.
p. 20 / Section 8	Other Issues: Updated description of the evolving nature of data across semi-independent mini-projects and confirmed the next update is planned for early 2026.

CHESS Consortium

Participant organisation name	Short name	Country
Masaryk University	MUNI	Czechia
University of Tartu	UTARTU	Estonia
Brno University of Technology	BUT	Czechia
Tallinn University of Technology	TalTech	Estonia
Cybernetica AS	CYBER	Estonia
Red Hat	RedHat	Czechia
Guardtime	Guardtime	Estonia
Estonian Information System Authority	RIA	Estonia
CyberSecurity Hub	CSH	Czechia
National Cyber and Information Security Agency	NCISA	Czechia
South Moravian Innovation Centre	JIC	Czechia
Estonian Association of Information Technology and Telecommunications	ITL	Estonia

Abbreviations

CA – Challenge Area
CHESS – Cyber-security Excellence Hub in Estonia and South Moravia
DMP – Data Managemet Plan
ICT – information and communication technology
KPI – key performance indicator
NGO – non-governmental organisation
R&I – Research and Innovation
TA – target audience
WP – work package
CVEs – Common Vulnerabilities and Exposures
MPC – multiparty computation
F4SLE – Framework for Security Level Evaluation

Table of Contents

1 Introduction

This document serves as an update to Deliverable *D5.2 Initial Data Management Plan*, submitted on 30 June 2023. It reflects the ongoing evolution of the CHESS project's research activities and data management practices.

The CHESS project builds around six cybersecurity Challenge Areas (CAs): Internet of Secure Things, Security Certification, Verification of Trustworthy Software, Security Preservation in Blockchain, Post-Quantum Cryptography, and Human-Centric Aspects of Cybersecurity.

Each Challenge Area operates with a high degree of autonomy, launching and conducting small-scale research projects based on defined research priorities. At the start of the project, each CA launched two to three such initiatives ("mini-projects"), resulting in a total of fifteen at the project's outset. The initial DMP focused on the datasets and management strategies relevant to these early-stage efforts. As the project has progressed, the scope and composition of small-scale projects have evolved. Some projects have concluded, while new ones have been initiated in response to emerging research directions. As of this update, seven mini-projects have been completed and nine new ones have been launched. These developments are regularly reviewed through an annual re-evaluation process, where each CA reports on progress and outlines plans for the upcoming year, including changes in focus or the proposal of new initiatives.

Since the beginning of the project, there have been no major changes to the data management practices outlined in the Initial Data Management Plan (DMP). The types and formats of data remain the same, and we do not anticipate any substantial increase in the volume of data to be generated or re-used. We continue to use the same repositories as previously specified, and our commitment to making data openly available remains unchanged. No new data outputs have been generated beyond those described in the Initial DMP. Minor updates have been made to reflect the current project status and to ensure continued alignment with best practices.

The CAs not only serve as the focal point for research but also contribute to all work packages, ensuring cross-cutting coordination and shared learning, including the area of data management. Throughout the project, we monitor the status and needs of each research effort to ensure proper handling of data in line with FAIR (Findable, Accessible, Interoperable, and Reusable) principles.

This updated DMP outlines the tools, workflows, and governance structures the consortium uses to manage data effectively. It builds on the foundation of the initial plan, while incorporating new developments to ensure alignment with the evolving objectives of the project.

1.1 CHESS in Short[1]:

The Cybersecurity Excellence Hub in Estonia and South Moravia (CHESS) brings together leading R&I institutions in both regions to build connected innovation ecosystems to address one of the most important issues confronting Europe today: cybersecurity.

The CHESS Hub has been designed specifically to meet the cybersecurity challenges, building two linked Hubs in Estonia and South Moravia (Czech Republic), two regions that have already become important centres of IT industry and research. We develop a joint strategy closely aligned with the priorities set by the H2020 flagship cybersecurity pilot projects.

[1] Extracts from the project website https://chess-eu.cs.ut.ee/

Our place-based ecosystems identify research and training needs and execute small pilot research projects as proofs of concept or to otherwise validate technologies or their business models. We seek to bridge the research-innovation gap by connecting fundamental researchers with economic and societal exploitation. CHESS seeks to leverage local and regional resources to support the development and sustainability of the ecosystem.

Project Objectives:

- Develop a cross-border joint cybersecurity research and innovation (R&I) strategy aligned with Czech and Estonian smart specialisation strategies and Europe's digital society and cybersecurity goals.
- Apply the strategy in six focus areas of cybersecurity, i.e., six Challenge Areas:
  - Internet of Safe Things
  - Security Certification
  - Verification of Trustworthy Software
  - Security Preservation in Blockchain Technology
  - Post-Quantum Cryptography
  - Human-Centric Aspects of Security
- Initiate at least 12 small-scale R&I projects consolidating academia-business linkages, demonstrate the validity of ideas, and provide evidence to obtain additional investments.
- Develop a training strategy for both regions to increase cross-border/sectoral cooperation and skills around the six priority areas.
- Raise visibility, citizen engagement, technology transfer, entrepreneurship training, staff exchange, and mutual learning in Cyber-Security.

1.2 Organisation of the CHESS Project:

The CHESS project is organised in 5 WPs:

WP1	In WP1, we conducted an analysis of the current state of the Cybersecurity Ecosystem in Estonia and South Moravia. The six CHESS Thematic Working Groups contributed to mapping the capacities, expertise, common interests, and needs in both regions. Based on the outcomes of this mapping exercise, the consortium developed a strategy to enhance the excellence and impact of cybersecurity research and innovation (R&I) in South Moravia and Estonia. We relied on data collected through structured questionnaires, interviews with CA leaders and co-leaders, and extensive desk research. We examined national and EU cybersecurity strategies, including the Estonian and South Moravian smart specialisation strategies, as well as major EU Horizon 2020 flagship projects (CONCORDIA, SPARTA, CyberSec4Europe, and ECHO). The work completed within this work package has resulted in strategic documents, specifically public project deliverables. These deliverables are published on the CHESS website, and those already approved by the European Commission are also available on https://zenodo.org/communities/chess.
WP2	In this WP, we organise training and skills-building activities for ecosystem actors and provide excellent training open to global audiences. The project partners create *training materials, presentations from workshops/seminars, and video recordings,* most of which are and will remain publicly available on the CHESS website. Some training might have limited access, e.g., specialised training for government partners.

WP3	Small-scale R&I actions are implemented under WP3 to support cross-regional integration and strategy development across CHESS six diverse challenge areas. Each mini project is led by different partners and addresses a diverse range of topics. The first report on these R&I actions, *D3.1 Mid-term Evaluation Report of CHESS R&I Activities*, has been published on the CHESS website. We plan to upload it to Zenodo as soon as it receives approval from the European Commission. Activities within WP3 result in a variety of outputs, including publications, Bachelor and Master theses, open-source implementations, open datasets, and joint presentations of results. Notably, some research efforts initiated in WP3 have led to successful new project proposals funded under Cluster 3 of Horizon Europe. These follow-up initiatives are coordinated by CHESS consortium partners. Data management will remain a key focus in these new projects, ensuring continuity in data standards, FAIR principles, and repository use. Where appropriate, data produced in WP3 will be re-used, extended, or referenced in the follow-up projects, contributing to long-term impact and sustainability of research outputs.
WP4	Dissemination, Exploitation, and Communication strategies are described in *D4.1 Dissemination, Exploitation and Communication Plan* of the CHESS project.
WP5	MUNI, responsible for WP5 Project Management, has established an electronic project monitoring workspace to keep all documents common to the project available to all partners. This space serves as an online working space for the partners and storage of all internal documents.

1.3 CHESS Challenge Areas

CHESS Challenge Areas (CA) are the cornerstones of the CHESS project activities. Most of the work done is on the level of individual Challenge Areas. Thematic working groups from each Challenge Area contribute to Strategy Development (WP1), Skills development actions and training (WP2), Research and Innovation (WP3) and Dissemination and communication (WP4). The following section provides a brief overview of all pilot research initiatives, including those launched at the start of the project that were reflected in the Initial Data Management Plan. This update also includes the current research activities being conducted within each of the six Challenge Areas.

CA1: Internet of Secure Things

Strategic Priorities

- Promote effective approaches that public and private organisations can take to support transition to and securely manage of IoST systems. Develop, validate and deploy IoST systems in various sectors, such as transportation.
- Improve security of IoST systems with advanced technologies.

Pilot research:

- Empirical Research on Security and Privacy Management in Intelligent Infrastructure Systems.
- Analysis of Security-Aware and Privacy-Preserving Smart Parking Solutions.
- Secure and Privacy-Preserving Access to Sharing Vehicles in Smart Cities.

Current Research:

- Security Risk Management in Automated Systems and Technology.
- Security and Privacy in Teleoperated Systems.
- Secure and Privacy-Preserving Access to Sharing Vehicles in Smart Cities (continuing pilot research).

CA2: Security Certification

Strategic priorities

- Develop lightweight and automated (re)certification processes to ensure scalability.
- Explain vulnerabilities in certified devices by structuring certification documents that could be easily (deterministically) processed in an automated fashion to enable linkage of certification data to new knowledge regarding vulnerabilities within certified devices (CVEs, etc.).
- Develop methods of cybersecurity certification and deployment that ensure all layers and threats are correctly weighted. Cross-referencing certified items to vulnerability databases, like common vulnerabilities and exposures (CVEs).
- Develop security certification labels for selected devices, software and organisations that provide a simple and unambiguous depiction of the level(s) of the security being certified.

Pilot Research

- Enriching Certification Report Analysis with other Open-Source Intelligence
- Testing and improving a Method for Evaluating Organisations' Information Security

Current Research

- Enriching Certification Report Analysis with other Open-Source Intelligence (continuing pilot research).
- Testing and improving a Method for Evaluating Organisations' Information Security (continuing).
- Developing Common Criteria Security Target for multiparty computing platforms enabling privacy-sensitive data processing.
- Verifying the certification claims of full disk encryption systems.

CA3: Verification of Trustworthy Software

Strategic Priorities

- Make use of Program Analysis Techniques to improve Software Development.
- Develop a theory of composable cybersecurity protocols to offer visual accounts of organisational cybersecurity protocols understandable to non-experts.
- Identify practically motivated challenges for basic research not yet covered by existing methodologies.

Pilot Research

- Development of Theory and Tool Support for Cybersecurity Protocols.
- Emerging problems in formal methods.

Current Research

- Continuing in all pilot research initiatives.

CA4: Security Preservation in Blockchain

Strategic Priorities

- Illustrate the state-of-the-art use of blockchain in vehicular communication environment.
- Develop building blocks for hardware wallets with multiparty computation (MPC).
- Investigate privacy enhancing technologies used in Bitcoin blockchain.
- Demonstrate that blockchain can be used to manage traffic signals for emergency vehicles.

Pilot Research

- Secure consensus for Intelligent Vehicular Communication.
- Emergency Information transmission using blockchain in Intelligent Vehicular Communication.
- Blockchain-related operation protected by cryptographic hardware with MPC.

Current Research

- Secure Information Transmission in Intelligent Vehicles.
- Privacy of blockchain transactions.
- Methods for More Compact and Secure Blockchains.

CA5: Post-Quantum Cryptography

Strategic Priority

- Evaluate the current state and practical applicability of post-quantum technologies.
- Assess usability & market viability of information security products based on post-quantum algorithms.

Pilot Research

- Evaluation of classical asymmetric algorithms (signatures, Diffie-Helman Key exchange, etc.) replaced by post-quantum algorithms (such as CRYSTALS-Dilithium and Kyber).
- Update and test the UXP data exchange layer (an electronic voting solution) with PQ.
- Post-Quantum Communication Infrastructure Pilot - establishing a fully functional quantum-safe communication channel between Czechia and Estonia.

Current Research

- Aspects of Transition to Post-Quantum Technologies.
- Post-Quantum Cryptography on Constrained Devices.
- Post-Quantum Communication Infrastructure Pilot (continuing).

CA6: Human-Centric Aspects of Security

Strategic Priority

- Improve cybersecurity training through hands-on and tabletop exercises.
- Improve usability of cybersecurity solutions for ICT professionals.

Pilot Research

- Evaluate Automated Feedback upgrade to KYPO Cyber Range Platform (an open-source interactive learning environment for hands-on cybersecurity training).
- Identify and address gaps in the usability of penetration testing reports among ICT professionals.

Current Research

- Design, delivery, and assessment of tabletop exercises.
- Improving the usability of penetration testing reports (continuing pilot initiative).

1.4 Data Management Principles

- The Coordinator cooperates closely with Masaryk University's relevant professional services, i.e., the MUNI Open Science team, Open Science Methodologist at the Faculty of Informatics and MUNI Personal Data Protection Office.
- Personal data processing in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). This will apply to all surveys, questionnaires or interviews (e.g., mapping in WP1) to safeguard participants' anonymity. More information on standard procedures for processing personal data at Masaryk University can be found here: https://www.muni.cz/en/about-us/official-notice-board/personal-data-protection
- We work in compliance with MUNI Open Science Strategy 2022–2028: https://openscience.muni.cz/media/3477939/en-a-strategicka-cast-strategie-open-science-mu-2022-2028-uprprebal.pdf
- The Data Management Plan (DMP) will be updated once a year or whenever a significant change may occur.
- Data Management is on the agenda of Strategy Board meetings (is regularly reviewed).
- CYBER is ISO/IEC 27001 and ISO 9001 certified. As such, CYBER has an ISMS (Information security management system) that dictates the rules for data management, among other things. Hence, the users have to either work on data kept in the internal servers with nightly backups or back up their computers in the internal back-up server.

2 Data Summary

2.1 Re-Use of Existing Data

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.

Challenge Areas re-use different types of existing data. For open-source implementation (e.g., prototype, demonstration), we use open-source code. We use existing open-source projects in the form of software and libraries. We use these components to build post-quantum systems (in CA5 Post-Quantum Cryptography). We also use Common Criteria and FIPS 140 certification documents and accompanying documentation (all public) and NIST National Vulnerability Database. In CA6, we used public reports on pentesting in the mini project focused on improving the usability of penetration testing reports. When preparing the Strategy within WP1, we worked with reports and policy documents published by various authorities and institutions. We also benefit from different training materials produced in training projects or courses when preparing training events within WP2.

2.2 Data Types and Formats

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

The CHESS project works only with well-known standard formats. Publications are typically PDF files that might be supported by some data. Master theses (textual, graphical) may include source code in case of prototypes and demonstrations. Regarding open-source implementations, the consortium works with repositories with a Docker container/GitLab repository, source code files, and related documentation. Survey results and benchmark results are in tabular format. Other formats used will include PPTX, JSON, text documents and video format. Most of the Challenge Areas generate data mainly in the form of publications and software. The re-used data then includes predominantly software code.

2.3 Purpose of Re-Used/Generated Data

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

Based on the information we collected during the mapping/preparation of the SWOT analysis in WP1, we identified and enabled actions that aim to reinforce R&I excellence in the regions and promote the impact of cybersecurity R&I in diverse sectors of society and the economy. WP2 and data generated (training materials, video recordings, presentations) aim to increase the needed skills of the relevant stakeholders and target groups on cybersecurity issues. Generated publications support the project's key objective of cybersecurity education and raising awareness. The purpose of re-used or generated data within the CHESS project is project demonstrations, scientific research, and results of the small-scale projects that can be further used by different stakeholders. For example, in one of the activities, we create novel data about relations between standardisation documents and vulnerabilities. In another activity, we collect survey responses and provide their mutual comparisons.

2.4 Size of Data

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

We did not experience and do not expect any significant size of data to be generated or re-used within the projects (GBs). In some of the Challenge Areas, no datasets are planned to be produced, only publications and software. Therefore, there are no special requirements for data storage within the project.

2.5 Origin of Data

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

Most of the data is publicly available (e.g., open-source data in the form of publications and code). Part of the data is collected from survey responders, e.g., data collected from organisations. For Master theses/publications/presentations, we use literature analysis, analytical research, and validation of the research results. Project partners develop open-source implementations using analytical research and validation of research results.

2.6 Data Utility Outside of the Project

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

The data generated within the project is useful to the scientific community, companies, the public, project partners, non-governmental institutions, security experts and managers. All publications aim to inform and educate actors from all four ecosystems (quadruple helix). Nearly all software(s) shall be reusable for future research and allow for repeatability and verification of our research.

3 Fair Data

3.1 Making Data Findable, Including Provisions for Metadata

3.1.1 Persistent Identifiers (PIDs)

Will data be identified by a persistent identifier?

Data generated within the project is identified by a persistent identifier. Publications are assigned standard identifiers, mostly DOI. Code is assigned identifiers given by the repositories, mainly GitHub, GitLab, etc. Some of the open-source implementations will be identified by Project ID on GitHub. Theses are published in repositories of the relevant institution (e.g., MUNI theses are published in the nationwide repository and are assigned persistent URLs). Data from one of the mini project (sec-certs in CA2 https://sec-certs.org/) is available with a persistent identifier in incremental versions. Statistics and results of analyses from another mini project (Framework for Security Level Evaluation F4SLE survey) are available in the form of publications (with a persistent identifier) and presentations. We also plan to use Zenodo communities to ensure all our research outputs have a persistent identifier, persistent storage, and a common collection gathering all outputs of the project.

3.1.2 Metadata and Their Standards

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Most metadata are made available through the project deliverables and open-access publications, as well as through the project dissemination channels (e.g., project website and social media accounts). Standard metadata collected and published by scientific databases and code repositories will be provided. For all results deposited in Zenodo, it is possible to export their metadata in standard formats such as MARCXML, Dublin Core and DataCite Metadata Schema (by means of the repository itself). The type of metadata depends on the specific activity and project output. For example, in one activity, metadata about relations of standards and other documents examined is the actual output of the activity:

- Protection profile (PP): PP Title, PP Version, the target of evaluation, Evaluation Assurance Level, CC Version, PP Author.

Some of the other metadata include:

- PhD/Master/Bachelor theses: author's name, abstract, graduation thesis language, graduation thesis type, supervisor(s), defence year, keywords.
- Publications: depends on the publisher.
- Presentations and video demonstration: author, title, short abstract, keywords.
- Benchmarks: author, title, explanation of dataset structure.
- Software code with accompanying data.
- Metadata related to project events. Some of our events have separate webpages.

The issue of metadata and its standards will be further specified in the next version of DMP depending on the progress and research focus of individual Challenge Areas.

3.1.3 Keywords

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

Search keywords are provided in the metadata. Target publication venues are open-access peer-reviewed journals and conferences due to vast dissemination opportunities. The publication venues are easy to access online and offer a simple keyword, author, or DOI (digital object identifier) search through their homepages or publication search engines such as "sciencedirect.com" or "scopus.com". In addition, references are made via the project's official website. The keywords are provided for Master theses, publications, video demonstrations, benchmark results and code.

3.1.4 Providing Metadata for Indexing

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

Publications (and relevant metadata) are indexed in scientific libraries (such as SCOPUS, WoS, ScienceDirect, IEEExplore) and software are indexed in open-source repositories (mainly GitHub and GitLab). For all results deposited in Zenodo, their metadata are exported via OAI-PMH and will allow for harvesting (by means of the repository itself).

3.2 Making Data Accessible

Repository:

Will the data be deposited in a trusted repository?

The consortium uses the official institutional repositories of the partners. For open-source implementation, these are institutional installations of Github/GitLab, e.g. https://gitlab.cs.ut.ee/, https://github.com/crocs-muni/. For theses, we use official university registries, e.g. https://is.muni.cz/thesis/ or https://thesis.cs.ut.ee/. Research and survey papers produced are published as open-access by taking up self-archiving rights for journals and conferences that have them, or if necessary, paying the open-access fees where self-archiving and or free open access is not possible. Some examples of free repositories we use include https://www.iacr.org/eprint/ or https://arxiv.org/. All final versions of publications are stored in accordance with journal/conference open access rules at several public repositories such as university libraries, ResearchGate or other open libraries. We use Zenodo communities (as mentioned above). The project website also serves as a repository for project outputs, such as publications, training materials, video recordings or presentations.

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

The repositories mentioned above have been used in the past with a positive experience and provide safe and long-lasting storage.

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

The publication and theses repositories ensure that data is assigned an identifier, i.e., DOI, ISSN, ISBN for publications. We use Zenodo communities for other research outputs, e.g., project deliverables.

Data:

Will all data be made openly available?

All data are made openly available, apart from survey data. The survey input data/raw data can be accessed only by the project partner(s) who provided or collected it, as the data contains sensitive organisational data. For certain coinjoin privacy experiments in the research focused on the privacy of blockchain transactions (CA4), we will not share the raw dataset in order not to decrease the privacy of other participants - only aggregated results will be published. In some of our mini-projects, we engage with companies to validate and refine our research proposals. In some cases, this collaboration is governed by Non-Disclosure Agreements (NDAs), which allow us to use the data for research purposes but prohibit public disclosure. As a result, only aggregated or deidentified results are published, and only when explicitly permitted by the respective companies.

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.

N/A

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

All repositories we use are accessible through HTTPS protocol.

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

In some cases, there are closed thesis defences, and there might be restrictions on access depending on the thesis license. Regarding surveys, only aggregated results will be provided to the project deliverables.

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

N/A

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

N/A

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

Yes, metadata is openly available.

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

All repositories we use provide long-term storage of data. Both data and metadata will remain available for at least ten years after the project end date or as long as needed.

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)?

Data generated within the project are in standard formats, and no additional software is needed to access or read the data. Publications are open-access, and software is published as open-source. The same applies to metadata.

3.3 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?

The consortium uses open data formats (common, well-known formats only) that are easily usable and will make it easy for anyone interested in working with the data. Publications and theses are available in standard formats; PDF is expected. Source code will be available through standard repositories in open-text format.

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 re-using, refining or extending them?

N/A

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

Our data includes qualified references to other data where appropriate. The open-source implementations will typically use (reference) the available libraries and other open-source implementations.

3.4 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.)? 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 re-use licenses, in line with the obligations set out in the Grant Agreement?

Our data is freely available in order to permit the widest re-use possible (apart from sensitive/raw data from surveys). All data will be available in trusted repositories (see above) and on the project website. Our data will be licensed using free licenses. Publications will be open access. For open-source implementations, readme files and related publications will be available.

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

Yes, as almost all of the results of the project will be accessible in respective/above mentioned repositories, with permissive licences, the data produced will be usable by third parties also after the end of the project.

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

Yes, where appropriate. For example, for open-source implementations, the provenance of the data will be documented through the commit tracking in Git Repository.

Describe all relevant data quality assurance processes.

We follow the relevant data quality assurance processes according to the applicable standards.

- Survey results: Checking the data entries on completeness and general correctness based on the open-access documents
- Theses are publicly defended at a university.
- Open-source implementations: The developed implementations will be tested against the functional requirements and the design goals.
- The publications are reviewed by the international committees.

4 Other Research Outputs

All project outputs are described in the sections above.

5 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?

There are no special costs expected for making data FAIR in our project. We use the standard services at the partner institutions. We need data specialists to support the consortium, but they are part of the standard service at the institutions, so they are not included in the project budget.

We expect only costs related to open-access conferences and journal fees for produced publications. The cost for OA publishing is included in the project budget.

Who will be responsible for data management in your project? How will long-term preservation be ensured?

Challenge Area Leads are responsible for data management within their area. They monitor progress and data management within individual mini projects. CA Leads are supported by the Project Coordinator and Open Science Team at Masaryk University and the institution they come from. All repositories we use and plan to use provide long-term storage of data. Both data and metadata will remain available for at least ten years after the project end date or as long as needed (see above).

6 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)?

All data is stored on the institutional storage infrastructures and automatically backed up.

As the CHESS consortium is focused on cybersecurity, and includes many data security and IT experts, we are confident that the project data is and will be well secured. For example, Cybersecurity Team of Masaryk University CSIRT-MU is the first certified cybersecurity team in the Czech Republic. Their main focus is to protect cyberspace at Masaryk University.

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

Yes, the data will be safely stored in trusted repositories mentioned above.

7 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).

According to the Grant Agreement, any research connected to the CHESS project will be carried out in compliance with fundamental ethical principles. If we identify an ethical issue, we will approach an ethical committee at a relevant institution to get approval and proceed according to standard procedures and regulations. We have already asked for approval from the Ethics Committee of the University of Tartu for the research regarding the usability of penetration testing reports. In the course of the research, some personal data will be collected, and Artificial intelligence (AI) technologies are used to create the content to be evaluated in the study. The outcome of this request will be received in the new review period of the project.

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

We follow GDPR standard procedures (see above).

8 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)?

The project is a cooperation of many semi-independent mini projects, where the types of data differ and keep changing as the project progresses. The Initial DMP described the data we were aware of at the beginning of the project. To reflect the changes within individual Challenge Areas, the consortium updates the DMP regularly. We plan the next update at the beginning of 2026 for the last year of the project.