Data Management Plan (DMP)

Behavioral Next Generation in Wireless Networks for Cyber Security (BEiNG WISE), CA22104

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

Data Management Plan (DMP)

Behavioral Next Generation in Wireless Networks for Cyber Security (BEiNG WISE), CA22104

On the other side, this connected world opens new breaches and creates new potential vulnerabilities for smart advanced cyber-attacks, namely attacks and offender relying on ML/AI and advanced wireless technology integration, to make their attack more effective and less detectable. If an increasing awareness by the users could help to contrast the security issues, it is not sufficient against the new generation of cyber-attacks. In this context, a drastic paradigm shift, putting human-being in the loop for the conception of novel and more effective cyber-security solutions, must be considered.

The always-connected world we are living in, gives us an unprecedented plethora of new advanced services and automated applications requiring, more and more, less human intervention due to the increased integration of Machine Learning (ML), Artificial Intelligence (AI) approaches and sophisticated emerging wireless technologies.

Human-beings have a double role in the cyber-connected world: as potential offender and potential victim. The focus of BEiNG-WISE will be on how these different human-being features can be combined with the advanced technological characteristics, in order to conceive non-conventional, responsible by design, cyber-security solutions accounting for both these factors. In this complex connected system, another fundamental aspect that needs to be accounted to, is the legal one, related to the conception of solutions that can be effectively employed in the real world. Also, legal aspects should be considered at the design stage. The Action relies on cross-domains expertise, ranging from cybersecurity, wireless communication technology, data science, sociology, psychology and law.

Working Group 1:
Cybersecurity in emerging wireless communications

Working Group 2:
A cybercrime perspective in wireless networks

Working Group 3:
Optimal Security approaches and their impact on the user

Working Group 4:
Human factors in wireless security networks

Working Group 5:
Legal Factors in Cybersecurity Wireless Systems: a vertical approach

Version changes

Data Management Plan (DMP)

Authors:	S. Kovačić, L. O'Toole
Version:	1.0
Date:	October 20, 2024
Status:	FINAL

Change history:

Version:	Date:	Authors:	Description:
1.0	October 20, 2024	S.K.	First final verson

Table of Contents

Executive summary

The always-connected world we are living in, gives us an unprecedented plethora of new advanced services and automated applications requiring, more and more, less human intervention due to the increased integration of Machine Learning (ML), Artificial Intelligence (AI) approaches and sophisticated emerging wireless technologies.

On the other side, this connected world opens new breaches and creates new potential vulnerabilities for smart advanced cyber-attacks, namely attacks and offenders relying on ML/AI and advanced wireless technology integration, to make their attack more effective and less detectable. If an increasing awareness by the users could help to contrast the security issues, it is not sufficient against the new generation of cyber-attacks. In this context, a drastic paradigm shift, putting human-being in the loop for the conception of novel and more effective cyber-security solutions, must be considered.

Human-beings have a double role in the cyber-connected world: as potential offender and potential victim. The focus of BEiNG-WISE will be on how these different human-being features can be combined with the advanced technological characteristics, to conceive non-conventional, responsible by design, cyber security solutions accounting for both these factors. In this complex connected system, another fundamental aspect that needs to be accounted to is the legal one, related to the conception of solutions that can be effectively employed in the real world. Also, legal aspects should be considered at the design stage. The Action relies on cross-domains expertise, ranging from cyber security, wireless communication technology, data science, sociology, psychology and law.

This report corresponds to Deliverable D4 - Data Management Plan (M12), 4.1.2. Description of Deliverables and Timeframe, of the BEiNG-WISE. This deliverable covers the specification of what data will be open, specifies the data the action will generate, whether and how it will be used or available for verification and reuse, and how it will be organized and preserved. It will be updated to reflect the latest state of action as it progresses.

Acronyms

AI	Artificial Intelligence
CDR	Common Data Representation
CERN	European Organization for Nuclear Research
DMP	Data Management Plan
DOI	Digital Object Identifier
DR	Disciplinary Repository
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HPDC	High Performance Data Center
HW	Hardware
IDL	Interface Description Language
IPR	Intellectual Property Rights
IR	Institutional Repository
ML	Machine Learning
MoU	Memorandum of Understanding
OpenAIRE	Open Access Infrastructure for Research in Europe
ORDP	Open Research Data Pilot
SW	Software

1. Introduction

The BEiNG-WISE initiative is a forward-thinking and ambitious action focused on advancing cybersecurity in wireless communication systems by integrating critical human factors into the development, design, and implementation of security protocols. This initiative recognizes that cybersecurity is a multifaceted challenge, one that extends beyond just technological innovation to include the complexities of human behavior, decision-making, and interactions with technology.

As wireless communication technologies continue to evolve rapidly, the methods and sophistication of cyber threats are advancing in parallel. BEiNG-WISE aims to not only address the technical challenges posed by these emerging threats but also bridge the critical gap between cutting-edge technological solutions and the nuanced understanding of human behavior that influences both the perpetration and prevention of cybercrimes.

By adopting a holistic approach, BEiNG-WISE seeks to enhance the resilience of wireless communication networks against a range of cyber threats while ensuring that human factors-such as user behavior, awareness, and response to security measures-are fully accounted for. This comprehensive approach promises to create more robust, adaptive, and user-friendly security frameworks, ultimately contributing to a safer and more secure digital environment for individuals and organizations alike.

COST action Overview

BEiNG-WISE stands for "**Behavioral Next Generation in Wireless Networks for Cyber Security**" and focuses on several key pillars to achieve its goals:

1. **Cyber security in Emerging Wireless Communication**: The action addresses the rapid growth of cyber security threats compared to mitigation measures. It explores the sustainability of current approaches in the face of advancing technologies like the Internet of Everything (IoE), Bring Your Own Device (BYOD), and the always-online paradigm. By examining these trends, BEiNG-WISE aims to outline clear research directions and paradigm shifts necessary for robust cyber security solutions.
2. **Cyber-Crime Perspective in Wireless Networks**: Understanding the motivations behind cyber-attacks is crucial. This aspect of the action delves into the psychological and social factors that drive individuals to engage in cyber-criminal activities. By comprehending these factors, the action aims to develop more effective preventive measures and resilient communication systems.
3. **Human Factors in Wireless Security**: Recognizing the importance of human factors in cyber security, BEiNG-WISE emphasizes the need for security solutions that are user-friendly and consider human behavior. This approach ensures that security measures do not overly complicate user experiences but instead enhance them while maintaining robust security.

Preparation of the Data Management Plan (DMP)

The Data Management Plan (DMP) is a critical component of BEiNG-WISE, ensuring that all data-related activities are conducted efficiently and securely. The preparation of the DMP involves several steps:

1. **Data Collection and Storage:** Identifying the types of data to be collected, including personal data, technical data, and behavioral data. Ensuring that data storage solutions are secure and compliant with relevant regulations.
2. **Data Access and Sharing:** Establishing protocols for data access to ensure that only authorized personnel can access sensitive information. This also includes guidelines for sharing data with external partners while maintaining security and confidentiality.

3. **Data Analysis and Utilization:** Outlining methods for analyzing the collected data to derive meaningful insights. This includes the use of advanced analytical tools and techniques to understand cyber threats and human factors in wireless security.
4. **Data Protection and Privacy:** Implementing measures to protect personal and sensitive data from unauthorized access and breaches. This involves adherence to legal and ethical standards for data protection.
5. **Documentation and Reporting:** Ensuring comprehensive documentation of all data-related processes and activities. Regular reporting to stakeholders to maintain transparency and accountability.
6. **Risk Management:** Identifying potential risks associated with data management and implementing contingency plans to mitigate these risks. This includes regular monitoring and updating of data management practices to address emerging threats and challenges.

By meticulously preparing and adhering to the Data Management Plan (DMP), BEiNG-WISE seeks to guarantee the integrity, security, and optimal utilization of data throughout the COST action lifecycle. DMP serves as a foundational element, ensuring that data is handled in a way that supports the action's overarching goal: developing cutting-edge, human-centric cybersecurity solutions tailored for wireless communication systems. This careful attention to data management not only enhances the technical robustness of the solutions but also ensures compliance with the best practices in privacy and data security, which are crucial in the modern digital landscape.

BEiNG-WISE represents a significant leap forward in the field of cybersecurity, particularly by integrating human factors into the design, implementation, and management of wireless security systems. Traditional cybersecurity measures often focus heavily on technical solutions, but BEiNG-WISE recognizes that human behaviors such as how individuals interact with technology, their awareness of security risks, and their responses to potential threats play an equally crucial role. The action's comprehensive approach addresses this interplay, targeting both the evolving nature of cyber threats and the vulnerabilities introduced by human factors.

By doing so, BEiNG-WISE aims to develop more adaptive, resilient, and secure wireless communication networks. This approach not only protects against immediate threats but also fosters a proactive, long-term cybersecurity posture. Ultimately, BEiNG-WISE seeks to pave the way for a future where wireless communication systems are not only technologically advanced but also designed with a deep understanding of the human behaviors that influence both security breaches and prevention strategies.

2. Data Cycle

Given the substantial volume and variety of data expected to be collected and generated throughout the BEiNG-WISE COST Action, ensuring its effective management, utilization, and long-term reusability is a top priority. This data will encompass a broad spectrum of formats and types, ranging from software (SW) libraries to hardware (HW) configurations, each of which plays a pivotal role in advancing the action's cybersecurity objectives.

To address the complexities involved in handling such diverse datasets, BEiNG-WISE has adopted a robust and comprehensive data lifecycle model. This model ensures that data is systematically managed at every stage, from its initial collection and storage to its analysis, dissemination, and eventual reuse by future researchers and developers. The lifecycle approach underscores the importance of data integrity, security, and accessibility, not only during the active phases of the COST Action but well beyond its timeline, enabling long-term contributions to the wider research community.

A key aspect of this strategy lies in the structured integration of Transversal Activities (Shown in the illustration: **Transversal Activities - TA3: Data Management and Figure 1**), which provide cross-cutting support for managing data from multiple disciplines and formats. These activities are designed to facilitate the alignment of data management practices with the FAIR principles ensuring that data is Findable, Accessible, Interoperable, and Reusable. By adhering to these best practices, BEiNG-WISE ensures that its datasets remain valuable, relevant, and applicable for future research endeavors, collaborations, and technological developments in wireless communication cybersecurity.

Moreover, this approach promotes transparency and collaboration within the action and beyond, fostering an open environment where the data can be shared and utilized by a broader network of stakeholders. Ultimately, the BEiNG-WISE initiative’s commitment to a comprehensive data lifecycle model not only supports the immediate goals of the action but also contributes to the sustainable growth and advancement of knowledge in the cybersecurity domain.

In the initial COST action phase, an initial version of the Data Management Plan (DMP) is formulated, outlining protocols for the handling of collected and generated data. These datasets undergo subsequent integration, preparation, and analysis, wherein they are meticulously described, categorized, and uniquely identified in accordance with COST action guidelines.

In a later phase of this COST action, the collected data undergoes a thorough review by BEiNG-WISE members to evaluate its significance and determine which datasets merit publication and open access availability. This process is critical to ensure that the most valuable and relevant datasets are made accessible to the broader research community, in line with BEiNG-WISE's strong commitment to open access principles. The importance of these considerations is highlighted in subsequent sections, as open access plays a central role in fostering transparency, collaboration, and the dissemination of knowledge generated by the COST action.

In addition to evaluating which datasets should be made publicly available, BEiNG-WISE establishes specific timelines for each dataset, clearly defining how long they will be retained for potential reuse by both COST action members and external researchers. These retention periods ensure that valuable data remains accessible for future research while balancing storage capacity and relevance over time.

Data storage options available to BEiNG-WISE include cloud infrastructure hosted by COST members, Institutional Repositories (IR), Discipline Repositories (DR) and self-managed High Performance Data Centers (HPDC). This flexible and scalable storage strategy ensures that data is securely stored, well

organized and easily accessible for reuse. This will enable seamless access to a wide range of stakeholders, including researchers, industry partners and external collaborators, ensuring that data remains both secure and highly accessible.

Since COST does not provide financial support to establish a centralized data storage solution for data management and sharing, BEiNG-WISE plans to use free and open access platforms such as GitHub and ZENODO for this purpose. GitHub offers robust version control and collaboration features, making it ideal for storing and sharing software, documentation, and code generated during COST action. ZENODO, an open access repository developed under the OpenAIRE program and hosted by CERN, enables the secure archiving and publication of datasets, ensuring compliance with the FAIR principles (findable, accessible, interoperable and reusable).

By leveraging these platforms, BEiNG-WISE not only addresses the need for efficient data management, but also promotes transparency, reproducibility and long-term availability of research results. These tools facilitate collaboration between action members and external researchers, while providing a cost-effective solution for archiving and disseminating action results to the wider scientific community.

A combination of flexible storage options ranging from member-hosted cloud infrastructure to free public repositories ensures that BEI-WISE can meet its current data management needs and long-term preservation goals. This strategic use of different storage solutions promotes data security, improves accessibility and supports the broader goals of knowledge sharing and collaboration within and outside the COST Action.

Throughout the entire data lifecycle, BEiNG-WISE places a strong emphasis on maintaining data quality through rigorous documentation, description, and evaluation processes. This ensures that all data is well-curated, properly described, and of the highest standard for reuse, whether by internal or external researchers. To keep pace with the evolving needs of the action, this data management document remains a living document subject to periodic updates and refinements as BEiNG-WISE progresses. These updates ensure that the action's data management practices remain aligned with the latest advancements in both technology and cybersecurity.

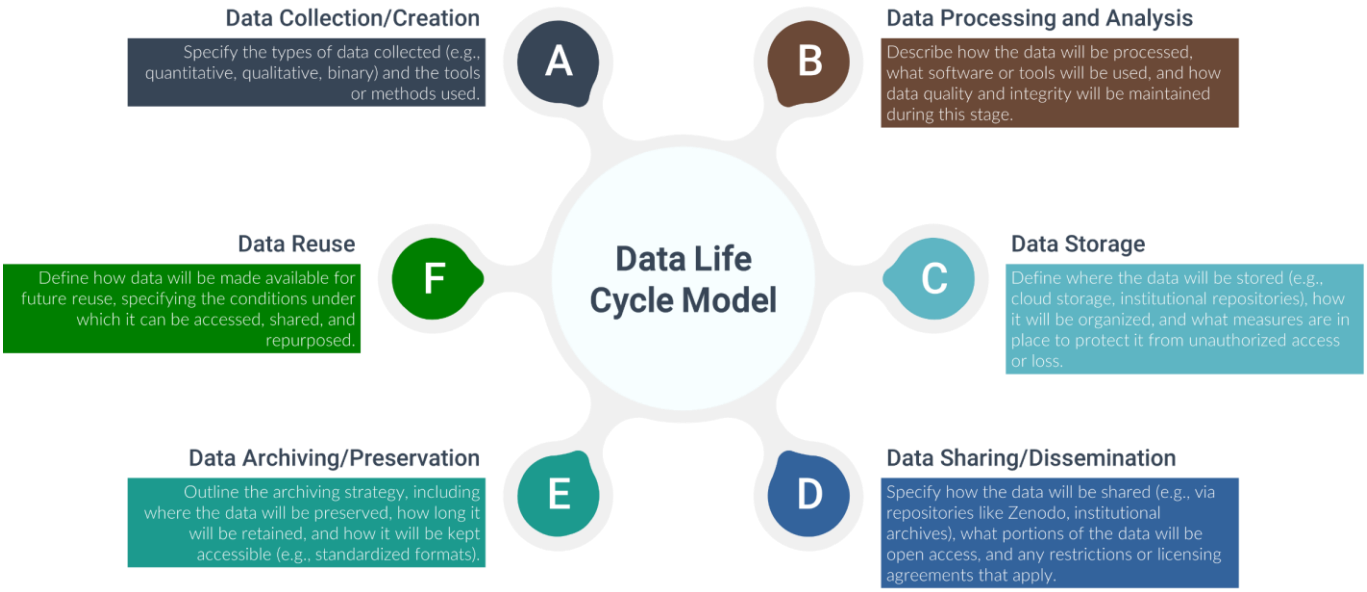

Figure 1. Data Life Cycle Model

3. Open Access Policy by BEiNG-WISE

In accordance with **Article 3.2.2. of the BEiNG-WISE MoU** (**Plan for Dissemination and/or Exploitation and Dialogue with the General Public or Policy**), members are required to promptly disseminate their findings in a publicly accessible format, while respecting any constraints imposed by intellectual property (IP) protection, security regulations, or other legitimate interests.

Following these guidelines, all action deliverables and data will be made publicly available, except for internal reports such as meeting notes or proprietary shared documents, which will remain private among BEiNG-WISE COST Action members. Research data will be deposited in publicly accessible repositories or published in scientific journals, ensuring broad access. Additionally, action deliverables will be cataloged in the publications section of the official BEiNG-WISE website.

Before public dissemination, BEiNG-WISE members will assess each dataset and generated knowledge to identify potential conflicts between open access and intellectual property rights (IPR) protection for commercialization purposes. Based on this assessment, decisions will be made regarding which information can be publicly disclosed and which must remain confidential. Members have two main strategies for safeguarding their findings: internal data preservation with selective external sharing after IPR measures, or pursuing patent protection for specific action outputs, followed by public disclosure once patent registration is secured.

These principles are outlined in **Article 3.2.2. of the BEiNG-WISE MoU**, which mandates that all beneficiaries diligently protect their results, considering factors such as commercial potential, the interests of other beneficiaries, and any relevant legitimate interests.

4. Data Summary

4.1. Data Description

The primary aim of the BEiNG-WISE COST Action is to advance non-conventional cybersecurity solutions for wireless networks by seamlessly integrating both technical and social elements. This innovative approach represents a paradigm shift in cybersecurity, placing human factors at the forefront of system design. By recognizing human beings as both potential attackers (offenders) and victims within the cybersecurity landscape, BEiNG-WISE aims to create security systems that are not only technically sound but also sensitive to the behavioral nuances that influence cyber threats and defenses.

The Action focuses on identifying and detecting novel vulnerabilities within wireless systems and developing robust, reliable, and human-centered countermeasures. This endeavor requires the generation and utilization of large datasets that will underpin the behavioral models necessary for the new framework. These datasets may be derived from the collective contributions of BEiNG-WISE members or obtained from publicly available repositories. By leveraging a combination of proprietary and existing data, the Action ensures ample resources for the development, training, and validation of new cybersecurity techniques.

Transparency and accessibility are key principles of BEiNG-WISE. Future iterations of this document will provide comprehensive details on the datasets being used, ensuring that stakeholders across academia, industry, and other sectors can fully engage with the research outcomes. This open approach fosters collaboration, promotes accountability, and enhances the overall impact of COST action innovative solutions in wireless cybersecurity.

The datasets generated or intended to be generated will be added in this document and characterized as follows:

Dataset summary

- **Dataset ID**: unique identifier of the dataset.
- **Name**: name of the dataset.
- **Lead Partner**: partner responsible of the dataset, its maintenance, and the data contained within it. This includes ensuring compliance with the data management and ethics requirements of the action.
- **Work Group**: references the Work Group for which this dataset has been used or generated.
- **Work Group Task**: action task which this dataset has been used or generated.
- **Description**: complete description of the dataset.
- **Data types**: type of that it contains (C++/Python/Java source code, images, text, etc.)
- **Data format**: format of the data.
- **New / Existing data**: specifies if the data has been created within the BEiNG-WISE COST action or already existed.
- **Mechanisms for data generation**: specify how the data was generated.
- **Expected size of the data**: specifies the expected size of the data in Bytes.
- **Utilisation of the data**: specifies how the data is going to be used within the action.
- **Quality control procedures**: specify the procedures to guarantee the quality of the data (peer reviews, tests, validation procedures, etc.)
- **Type of access (Open / Restricted)**: specifies the data protection issues.
- **Ethical issues**: specify whether the data could raise any ethical issue.

Making data findable

- **Data dissemination**: specifies how the data will be available.
- **Metadata standard**: specifies the metadata used to easily find the data.
- **Type of associated metadata**: specifies the type of metadata used.
- **How will the data be findable**: specifies the mechanisms used to make the data findable.
- **Software required to use / read the data**: specifies if there is any software requirement for the user to read the data.

Storage of data

- **Data storage location (short-term)**: specifies where will be stored in the short term.
- **Data storage location (long-term)**: specifies where will be stored in the long term.
- **Storage media**: specifies the storage media to be used.
- **Data security provisions**: specify how data security is ensured, in case the data is not publicly accessible.
- **Expected size of the data**: specifies the expected size of the data in Bytes.
- **Person responsible**: specifies the person responsible for ensuring the availability of data in the short and long term.
- **Cost**: specifies the cost of data storage.

Reuse of data

- **Reuse of existing data**: if the data already existed, specifies how the data is going to be reused within the project.
- **Potential reuse of data**: specifies how the data is going to be reused.
- **How will data be reused**: specifies how the data will be used within the framework of the action and the possibility of its use by the same or other entities within or outside the action at the end of the action.
- **Type of access (Open / Restricted)**: specifies whether the data is freely accessible, or access is restricted to the consortium.
- **Type of IP / protection sought**: specifies whether the data requires intellectual property conditions to be applied to it.
- **License**: specifies the license that applies to the data (Open-Source license or more restrictive licenses).

Other Comments: in this section, the person responsible for the data may specify other types of conditions for the publication and use of the data, if any.

A comprehensive description of each generated dataset, including detailed information on all relevant parameters such as format, structure, source, and purpose, will be provided in the description of each individual dataset. This section will also include additional metadata, such as data collection methods, processing techniques, and any relevant usage constraints, ensuring that all datasets are well-documented for ease of reuse and understanding by third-party researchers.

5. FAIR Data

This section outlines the methodology used to ensure that data management within the BEiNG-WISE framework aligns with the FAIR data principles (Findable, Accessible, Interoperable, and Reusable). Some datasets may not initially meet these criteria due to specific limitations outlined in TA3 - Data Management, and these will be clearly flagged and identified. As these datasets progress toward public availability, the Data Management Plan (DMP) will be updated accordingly to reflect their growing compliance with FAIR principles.

Tasks such as TA 3.1, which defines the rules for data sharing and storage, and TA 3.2, which encourages and facilitates data sharing, play a crucial role in guiding this transition. These ongoing updates will ensure that all datasets are systematically aligned with FAIR standards, enhancing transparency, accessibility, and reusability for the broader research community and other stakeholders. By establishing a solid foundation for data sharing, BEiNG-WISE ensures that even datasets initially not in full compliance will evolve to meet established FAIR guidelines over time.

5.1. Making Data Findable

The BEiNG-WISE metadata serves as essential descriptors, facilitating the swift and efficient discovery of research materials relevant to the action by interested third-party researchers. By employing precise, comprehensive, and accurate metadata, datasets generated by the BEiNG-WISE COST Action will be easily discoverable by the wider research community. These metadata will cover crucial elements such as data type, quality, availability, authorship, content description, versioning, and more, all presented in standardized formats within publicly accessible BEiNG-WISE research data repositories.

In compliance with TA3 - Data Management of the BEiNG-WISE Memorandum of Understanding (MoU), bibliographic metadata will follow standardized formats, ensuring consistency and ease of access. These standardized formats will include the following key elements:

Publication (author(s), title, date of publication, publication venue):

- COST Action funding[1]
- COST action name, acronym and number
- Licensing terms
- Persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant
- Where applicable, (...) persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication

The Digital Object Identifier (DOI) will be the persistent and unique identifier of the action's publications in open data repositories.

[1] All joint publications, co-authored by at least two members from two different countries of the BEiNG-WISE should indicate the EU funding of the COST action by including the following statement: "This [insert document type] is supported by European Union's COST under MoU BEiNG-WISE (Behavioral Next Generation in Wireless Networks for Cyber Security)."

A naming convention will be used to classify and identify the research results. These names are composed according to the following rule:

BW_ XX_DS_ID_YY_ZZ

were

- **BW:** the first two letters of name of the COST action BEiNG-WISE
- **XX:** date of creation according to ISO 86014 format (YYYYMMDD)
- **DS:** dataset name. The name must not contain blank spaces
- **ID:** dataset identifier
- **YY:** data type
- **ZZ:** version number starting at v1.0. The rule applied to dataset versioning follows the same rule applied to software versioning, i.e. vMajor.Minor. A change to the Major version indicates a significant change in the dataset that may have repercussions on the intended use or context to which it applies.

To assist third-party researchers in locating datasets, comprehensive metadata will be developed specifically for the BEiNG-WISE datasets. This metadata will incorporate keywords that succinctly describe and define the data, enhancing searchability and ensuring that relevant datasets are easily accessible.

5.2. Making Data Openly Accessible

Datasets generated during the BEiNG-WISE action will initially be stored in a shared repository accessible to all members of the COST Action. This repository serves as a collaborative space for data processing, analysis, and evaluation prior to publication, ensuring that all datasets undergo thorough scrutiny. Members will have the opportunity to contribute insights and feedback, fostering a culture of collaboration and shared expertise. The final, stable versions of the datasets stored in the BEiNG-WISE members private repository will be thoroughly reviewed by COST members in accordance with the guidelines outlined in DMP. Once approved, these datasets will be uploaded to GitHub, which will serve as the primary platform for dataset storage and sharing. GitHub's robust version control and collaborative features ensure that datasets are not only securely maintained but also easily accessible to researchers, fostering transparency and enabling seamless data reuse. This rigorous evaluation process is designed to uphold the quality and integrity of the data, paving the way for its eventual release to the broader research community.

In cases where no measures are taken to protect intellectual property rights or patents, these datasets may be published in ZENODO, a general-purpose open-access repository developed as part of the Open Access Infrastructure for Research in Europe (OpenAIRE[2]) program, managed by CERN. Additionally, arXiv[3] a free open-access archive and distribution service for academic articles will be considered for other types of publications, particularly as materials on this platform do not undergo peer review.

Data will be stored and published in standard digital formats that can be easily accessed using common software available on most computers. While not all types of metadata for publication are explicitly defined, we can establish some general formats, including:

- **PDF** format for documents and publications.

[2] OpenAIRE: https://www.openaire.eu/
[3] arXiv: https://arxiv.org

- **PNG, SVG or GIF** format prefer higher quality images, although they can be published in other formats such as JPG, in case image compression is required, or PDF format.
- **Audio** files will be released in MP3 (MPEG-1 Audio Layer III) format as a lossy audio format, and WAV as an uncompressed audio format.

- For **video** files, there are several options depending on the use case:
- ✓ **WEBM** format for videos to be published on the BEiNG-WISE website or used in other activities that require video streaming such as online lectures or conferences.
- ✓ **MP4 or AVI** format for high quality videos stored in the database.
- ✓ **WMV** format for compressed videos that require a smaller size, although this format is not compatible by default with Apple devices that do not have Windows Media Player software installed.
- **Interface Description Language** (IDL) files for generating CDR-compatible types.
- **Source code** files must be published in formats appropriate for the programming language. Some of these languages used are Python, C/C++ and CMake. Regardless of the use of these files, they can all be opened and read as plain text documents.
- Preferably **CSV** format for training, validation, and testing datasets used in the BEiNG-WISE framework. In case the data structure requires it, the JSON format can also be used for datasets.

The BEiNG-WISE action is also expected to generate binary data from experiments conducted within its framework. At this early stage, the specific software needed to process this binary data cannot be determined, but details will be outlined in future versions of this document. To facilitate open access to the data, two distribution methods will be employed: internal distribution, which is restricted to members of the BEiNG-WISE COST Action, and public distribution, which will be available to third-party researchers and the broader community interested in the subject.

For the first, i.e. internal distribution to action members, the following will be used:

- Proprietary data centers of the institutions cooperating on the COST action.
- Google Drive file sharing.
- Private part of the official website of the COST action BEiNG-WISE.

The following platforms will be used for open data access:

- ZENODO for the publication of data and metadata on the results and progress of research, development processes, implementation and testing.
- arXiv a free open-access archive and distribution service for academic articles will be considered for other types of publications, particularly as materials on this platform do not undergo peer review.
- GitHub for hosting all source codes that are defined as public and open to the user. This platform will also be used to develop a file sharing system where public action data will be included.
- YouTube for hosting presentations and speeches during events, conferences and workshops in which members of the COST action BEiNG-WISE participate, thereby giving greater exposure to the action.
- The BEiNG-WISE website.
- Other scientific journals are yet to be specified.

To manage access to private data, the institution hosting the data will grant access to researchers and action members on an individual basis, responding to each access request specifically. In contrast, public data is readily available without registration, and applicants do not need to undergo any authentication process to access it.

5.3. Making Data Interoperable

The previous section outlines the data formats used for all datasets generated by BEiNG-WISE, carefully selected to ensure seamless interoperability. All data generated in the action will adhere to the criteria defined in DMP, ensuring compatibility across different platforms and facilitating data exchange and reuse among researchers, institutions, organizations, and countries.

For all types of data, a standardized vocabulary will be used to describe data and metadata. BEiNG-WISE plans to adopt the DataCite metadata schema to standardize metadata, ensuring consistent and high-quality documentation. Additionally, keywords listed in the DMP will be used to organize, classify, and index data and metadata to improve visibility.

Given that BEiNG-WISE spans multiple fields of study including artificial intelligence, human-computer interaction, hardware acceleration, and sustainable computing a custom standardized vocabulary will be developed to ensure data interoperability across these disciplines. This vocabulary will be continuously updated as the action progresses, ensuring that it remains relevant and supports effective data sharing across all domains involved in the COST Action.

5.4. Increase Data Re-use

To maximize the reusability of BEiNG-WISE public data, Creative Commons licenses will be applied to protect the authorship of the generated datasets. These licenses grant copyright protection while ensuring proper attribution of the data and allowing third parties to copy, distribute, and use BEiNG-WISE results without the need for formal registration by COST Action members. Attribution-ShareAlike (CC BY-SA) and Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) licenses will be considered for BEiNG-WISE datasets. Both licenses permit third parties to distribute, modify, adapt, and build upon the data, provided they credit the original authors. These licenses also require that any derivative works be shared under the same terms as the original. The primary distinction between the two licenses is that CC BY-SA allows for commercial use of the data, while CC BY-NC-SA restricts it to non-commercial purposes.

BEiNG-WISE public data will be made available as soon as it has been reviewed by other COST Action members and a stable version is ready for submission. However, if the data is subject to intellectual property (IP) or patent protection, a journal or relevant entity may impose an embargo period. In such cases, BEiNG-WISE members will ensure that the embargo does not exceed six months.

As outlined in the BEiNG-WISE Memorandum of Understanding (MoU), data access requests will remain available for reuse for at least one year after the completion of the COST Action. Each collaborator is responsible for their own results and must provide access to the data and metadata for other users and related entities up to one year after the action concludes.

Data quality assurance actions are dictated and implemented by the institution that generates the data as the data owner. **These actions ensure that data sets**:

a) do not contain incorrect or outdated records,
b) provide the necessary metadata for a complete description of the data,
c) follow the conventions specified in this document for compliance with the FAIR data policy, and
d) are regularly updated during the reuse period.

6. Resource Management

This section offers a forecast of the expenses associated with FAIR data compliance within BEiNG-WISE. It delineates the financial coverage for these costs, outlines the accountability for data management, and addresses the strategy for the long-term preservation of resources.

Initial expenses associated with adherence to the **FAIR data policy include**:

- Publication costs for scientific articles in Gold Access journals, to be covered by the article's owning organization.
- Operational expenses for website maintenance, yet to be determined.
- Creation and upkeep expenses for the Github repository hosting the source code developed at BEiNG-WISE. Since all repositories will eventually be made public, this service is provided at no cost.
- Establishment and maintenance costs for private repositories held by each partner to safeguard their results. The specific cost remains undetermined and will be the responsibility of each partner.
- Publication costs on ZENODO and arXiv are both free of charge.
- No cost associated with Open-Source licenses.
- No cost associated with Creative Commons copyright licenses.

As outlined in the BEiNG-WISE Memorandum of Understanding (MoU), each member of the COST action retains full ownership and is responsible for managing the results they have generated throughout the action. Members are also obligated to take appropriate actions to protect and exploit these results where applicable. In cases where two or more collaborators jointly produce results, and it is impractical to distinguish individual contributions, joint ownership of the results will be granted to all contributors. This ensures that all parties involved share the rights and responsibilities associated with the use, protection, and commercialization of these jointly produced outcomes.

Moreover, members are required to provide access to their data and results to other beneficiaries of the BEiNG-WISE COST Action. This facilitates collaboration and enables other partners to build upon existing findings. Access will also be extended to affiliated entities for up to one year following the conclusion of the BEiNG-WISE action, ensuring that collaborative research and exploitation of the results can continue even after the action officially ends.

During the action, research findings, data, and associated metadata will be shared among collaborators using the repository, administered by GitHub. This centralized repository ensures secure and efficient data sharing while maintaining proper access controls. In addition to using the shared repository, each collaborator is required to maintain a backup of their results in their respective institutional repositories, such as Institutional Repositories (IRs), Disciplinary Repositories (DRs), or High-Performance Data Centers (HPDCs). These multiple layers of storage provide redundancy and security, minimizing the risk of data loss.

In accordance with the MoU stipulations, GitHub is also responsible for the publication, maintenance, and facilitation of the reuse of the specified action deliverables. This ensures that all published results remain accessible for at least one year following the conclusion of the COST Action. By overseeing the long-term availability of these outputs, eProsima guarantees that researchers and external stakeholders will have continued access to important findings, enabling further exploitation, validation, and reuse of the data and results generated by BEiNG-WISE.

This collaborative framework, clearly outlined in the MoU, ensures that data ownership, access rights, and dissemination obligations are fully respected, promoting transparency, continued innovation, and responsible management of the action's outcomes well beyond its completion.

7. Data Security

Data security within the BEiNG-WISE initiative is a top priority, achieved through the implementation of robust access controls that ensure only authorized personnel can access sensitive data. These controls are designed to prevent unauthorized access and misuse, maintaining the confidentiality of all information within the action. To further safeguard sensitive data, all communications and transmissions between COST Action members are encrypted. Passwords and other sensitive credentials are shared securely through direct communication methods, avoiding potentially insecure channels like email.

Each member organization, which stores sensitive data, also plans to implement strict internal security measures for its data centers. Data must be stored in at least two geographically separate locations, reducing the risk of loss due to hardware failure, natural disasters, or other unforeseen events. This redundancy is crucial to ensure that critical data remains intact and accessible even in the event of localized disruption. Additionally, a mandatory bi-weekly backup of all generated data is enforced, reinforcing data integrity and ensuring continuity in data availability.

Upon the conclusion of the COST action, the responsibility for maintaining the security of the preserved data will transition to the individual members who own the data. These members will be required to uphold the same high standards of data protection that were in place during the action, ensuring that sensitive information remains secure well beyond the action's lifespan. This long-term approach guarantees the confidentiality and integrity of the data, preventing unauthorized access or data breaches in the future.

By employing a combination of strong encryption, redundancy in storage, regular backups, and clear post-action responsibilities, BEiNG-WISE ensures that all sensitive data is handled with the utmost care and remains secure throughout its lifecycle, from collection to preservation.

8. Ethical aspects

This section deals with the ethical and legal aspects related to the collection, handling and processing of personal data arising from interactions with people. In the context of BEiNG-WISE, special attention will be paid to mechanisms for storing, sharing, preserving and protecting the identity of individuals and organizations participating in various research projects resulting from BEiNG-WISE. When any data set contains confidential personal data or sensitive information, this data will be subject to assessment by the organization owning the data and members of BEiNG-WISE to prevent the release of any personal data that identifies the participants, which could be misused by third parties.

8.1. Ethical Self-Assessment

BEiNG-WISE ensures secure access of data for all its members by restricting access exclusively to researchers and developers actively collaborating on the action. Access rights to the shared data repository are granted and managed individually for each collaborator, ensuring that only authorized personnel can retrieve or modify action data. The access policy is centrally managed by companies like GitHub which oversees permissions for the shared repository. Confidential data transmitted between COST action members is encrypted, and passwords are communicated directly through secure channels, avoiding email for enhanced security.

Each collaborating organization is responsible for securing its data centers by storing data in at least two separate locations to mitigate the risk of loss. Additionally, a bi-weekly backup of all generated data is required to further prevent data loss. Once the action concludes, the ongoing security of preserved data will be the responsibility of the partners who own the data.

8.2. Data Collection and Anonymization Process

All data collected for this action and its scope must meet several requirements in order to be suitable for use within BEiNG-WISE:

- Under no circumstances will the data of minors or persons who cannot give consent be collected or used.
- Collected data must not be sensitive or personal data under any circumstances.

All such data will be fully anonymized before use. This will not imply a loss of information for the action, as the intent of this data will always be designing considerations for BEiNG-WISE framework-human interaction. Therefore, anonymized data will have the same behavior as non-anonymized data.

Depending on the data collected, the requirements of the specific experiment and the consent of each experiment participant, anonymization processes will be applied. Complete anonymization of data will be understood as a process over the data collected, whereby the final data cannot be de-anonymized, that is, it will not be possible to obtain personal data from the person behind the specific data.

8.3. Collection of Personal Sensitive Data

Members of the BEiNG-WISE COST Action will take proactive measures to avoid the collection of sensitive personal data. This commitment stems from a strong ethical obligation to protect individuals' privacy and ensure data integrity. Furthermore, none of the action members intend to solicit participation from children or individuals who are unable to provide informed consent. By prioritizing ethical research practices, the

BEiNG-WISE initiative aims to foster a safe and responsible environment for all participants while adhering to relevant legal and ethical guidelines.

8.4. Informed Consent

To collect data from participants involved in interviews, workshops, or any other BEiNG-WISE activities, individuals will be required to review and sign an Informed Consent Form prior to participation. This ensures that participants are fully informed about the purpose, process, and potential use of the data being collected.

9. Other issues

There are no issues to report at this stage of the BEiNG-WISE.

10. Datasets

This section provides a detailed definition for each of the datasets used within the context of the BEiNG-WISE. Since the COST action is at a relatively early stage, not all datasets are yet defined. Future versions of this document will update this section with newly generated datasets and possible updates on the dataset already described.

References

[1] OpenAIRE. How to comply with Horizon Europe mandate for Research Data Management. url: https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm (visited on Mar. 1, 2023).

[2] European Commission. Open access and Data management. url: https://ec.europa.eu/research/participants / docs / h2020 - funding - guide / cross - cutting - issues / open - access - data - management/data-management_en.htm (visited on Feb. 11, 2021).

[3] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: Scientific Data 3.1 (Mar. 2016), p. 160018. issn: 2052-4463. doi: 10.1038/sdata.2016.18. url: https://doi.org/10.1038/sdata.2016.18.

[4] European Union. 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). Apr. 2016. url: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[5] Line Pouchard. “Revisiting the Data Lifecycle with Big Data Curation”. In: International Journal of Digital Curation 10 (June 2015). doi: 10.2218/ijdc.v10i2.342.