Deliverable 1.3

Data Management Plan

Technical References

Document version	:	1.0
Submission Date	:	28/02/2021
Dissemination Level	:	PU
Contribution to	:	WP1-Project Management
Document Owner	:	UNINOVA
File Name	:	Data Management Plan
Revision	:	3.0

Project Acronym	:	BIECO
Project Title	:	Building Trust in Ecosystem and Ecosystem Components
Grant Agreement n.	:	952702
Call	:	H2020-SU-ICT-2018-2020
Project Duration	:	36 months, from 01/09/2020 to 31/08/2023
Website	:	https://www.bieco.org

This project has received funding from the European Union´s Horizon 2020 Research and Innovation Programme under Grand agreement No. 952702.

Revision History

REVISION	DATE	INVOLVED PARTNERS	DESCRIPTION
0.0	24.12.2020	UNI	Skeleton Creation
0.1	11.01.2021	UNI	Completed contents
1.0	25.01.2021	UNI	First internal completed version
1.1	30.01.2021	UNI	Review by internal reviewer, Ricardo Peres
1.2	03.02.2021	UNI	Implementing internal reviewers suggestion and update
1.3	05.02.2021	IFEVS	Review by Work Package Leader, Pietro Perlo
1.4	08.02.2021	UNI	Implementing reviewer suggestion and update
2.0	10.02.2021	UNI	Internal Final Version
2.1	21.02.2021	UMU	Review by external reviewer, Sara Matheu
2.2	22.02.2021	UNI	Implementing reviewer suggestion and update
2.3	23.02.2021	UMU	Review by external reviewer, Sara Matheu
2.4	23.02.2021	UNI	Implementing reviewer suggestion and update
2.5	24.02.2021	UTC	Review by external reviewer, Ovidiu Cosma
2.6	25.02.2021	UNI	Implementing reviewer suggestion and update
3.0	27.02.2021	UNI	Final Version (3th) of Data Management Plan

List of Contributors

Deliverable Creator(s): Sanaz Nikghadam-Hojjati (UNINOVA), José Barata (UNINOVA)

Reviewer(s): Pietro Perlo (IFEVS), Ricardo Peres (UNINOVA), Sara Matheu (UMU), Ovidiu Cosma (UTC)

BIECO project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952702

Acronyms

Acronym	Term
BIECO	Building Trust in Ecosystem and Ecosystem Components (Project Acronym)
CERN	The European Organization for Nuclear Research
CO	Confidential, only for members of the consortium (including the Commission Services)
CRIS	Current Research Information System
CERIF	Common European Research Information Format
D	Deliverable
DMP	Data Management Plan
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
ICT	Information and Communication Technology
IPR	Intellectual Property Right
OA	Open Access
ORD	Open Research Data
ORDP	Open Research Data Pilot
PSB	Project Steering Board
TL	Task Leader
USB	Universal Serial Bus
WP	Work Package
WPL	Work Package Leader
XML	eXtensible Markup Language

Executive Summary

This deliverable present the Data Management Plan (DMP) of BIECO project which is intended to be a 'living' document that will outline how the BIECO research data will be handled during and after the project, and so it will be reviewed and updated at regular intervals. It presents the data collected in the project and how the project will make this data FAIR (Findable, Accessible, Interoperable, and Reusable), in accordance with the concept of FAIR data management.

If follows the template of DMP provided by the EC in the *Guidelines* on FAIR Data Management in Horizon 2020, version 3.0, July 2016[1], and provides technical details on the data collected, as well as purpose for data collection, data utility and what, where and how data can be accessed and reused.

Project Summary

Nowadays most of the ICT solutions developed by companies require the integration or collaboration with other ICT components, which are typically developed by third parties. Even though this kind of procedures are key in order to maintain productivity and competitiveness, the fragmentation of the supply chain can pose a high risk regarding security, as in most of the cases there is no way to verify if these other solutions have vulnerabilities or if they have been built taking into account the best security practices.

In order to deal with these issues, it is important that companies make a change on their mindset, assuming an "untrusted by default" position. According to a recent study only 29% of IT business know that their ecosystem partners are compliant and resilient with regard to security. However, cybersecurity attacks have a high economic impact and it is not enough to rely only on trust. ICT components need to be able to provide verifiable guarantees regarding their security and privacy properties. It is also imperative to detect more accurately vulnerabilities from ICT components and understand how they can propagate over the supply chain and impact on ICT ecosystems. However, it is well known that most of the vulnerabilities can remain undetected for years, so it is necessary to provide advanced tools for guaranteeing resilience and also better mitigation strategies, as cybersecurity incidents will happen. Finally, it is necessary to expand the horizons of the current risk assessment and auditing processes, taking into account a much wider threat landscape. BIECO is a holistic framework that will provide these mechanisms in order to help companies to understand and manage the cybersecurity risks and threats they are subject to when they become part of the ICT supply chain. The framework, composed by a set of tools and methodologies, will address the challenges related to vulnerability management, resilience, and auditing of complex systems.

[1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Partners


Disclaimer

The publication reflects only the author´s view and the European Commission is not responsible for any use that may be made of the information it contains.

Table of Contents

List of Figures

List of Tables

1. Introduction

The Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and/or generated by the BIECO project. By BIECO being a H2020 project this DMP follows the Horizon 2020 DMP template that was designed to be applied to any Horizon 2020 project that produces, collects or process research data. The DMP aims at defining the management strategy of data generated during the project with the purpose to making research data FAIR.

This first DMP describes the data management principles and strategies, tools and BIECO data: data set, “Open Research Data Pilot” (ORDP) and BIECO Demonstrator that will be produced as part of the project activities and that are relevant to be included in the DMP. The consortium also targets open access when publishing papers and articles.

The DMP is a living document to be updated as the implementation of the project progresses and when significant changes occur, such as (but not limited to) new data, changes in consortium policies and changes in consortium composition and external factors. Therefore, the DMP will be reviewed and updated at regular intervals.

2. Data Summary

The BIECO DMP aims to provide a strategy for managing key data generated and collected during the project and optimize access to and re-use of research data. The main motivation of data collection/generation in BIECO project is to identify vulnerabilities across complex ICT systems. Some of the data that is collected (e.g. Task 3.2) and generated (e.g. Task 3.3) for this purpose is generally confidential and will not be open access. For example in Task 3.2 the collected data will be a combination of use cases dataset and public resource. The use cases dataset is confidential and could not be published for security reasons. Project partners should make sure that all possible data are made available for verification and re-use unless the task leader can justify why data cannot be made openly accessible.

All European Union-funded projects must try to disseminate as much information as possible and on top of that, the BIECO project was signed up to the ORDP which means that we are committed to giving open access to data generated during the project unless it goes against our legitimate interests. In this regard, the main purpose of the DMP is to ensure the accessibility and intelligibility of the data generated during the BIECO project in order to comply with the guidelines of the ORDP. Each data set created during the project will be assessed and categorized as open, embargo, or restricted by the owners of the content of the data set.

All the datasets, regardless of their categorization, will be stored in each of the participant entities' databases and in the Microsoft Teams folder created as an internal database of the partners. In addition, those categorized as open or embargo will be publicly shared (in the case of the embargo, after the embargo period is over) through the public section of the project website[2] and ZENODO[3].

ZENODO is an open access repository for all fields of science that allows uploading any kind of data file formats, which is recommended by the Open Access Infrastructure for Research in Europe (OpenAIRE).

2.1 Data Management Plan (DMP) Guiding Principles

The Data Management Plan of BIECO is realized within the Work Package 1.

The BIECO project DMP follow the principle of Open Access (OA) according to the Horizon 2020 guideline summarized in the Diagram presented in Figure 1.

[2] https://www.bieco.org/
[3] https://zenodo.org/

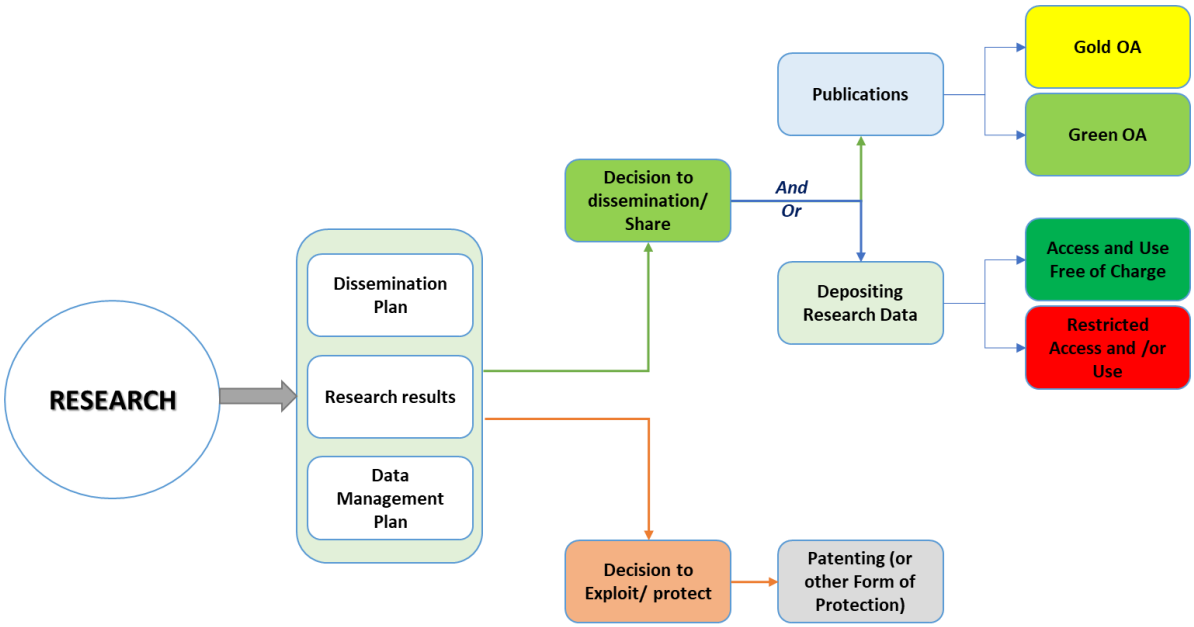

Figure 1 Open access to research data and publication decision diagram (from Guidelines to the Rules on Open Access to Scientific publications and Open Access to Research Data in Horizon 2020)

The others main principles of the BIECO project´s DPM are as following:

- This DMP has been prepared by taking into account the template of the "guideline of the Data Management in Horizon 2020"[4].
- The DMP is an official project Deliverable (D1.3) due in Month 6 (February 2021), but it will be updated throughout the project, and the final version of this document (D 1.4) will be available in month 32. By finalizing each BIECO document, such as deliverable, report, article, workshop report, conference abstract, the responsible partner required to complete Data Inventory Table (available in Annex 1) for the data that collected and generated in their document, after and deliver it to the UNI, for updating DMP. The first initial version will evolve depending on significant changes arising and periodic reviews at relevant reporting stages of the project.
- The consortium complies with the requirements of Regulation (EU) 2016/679 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, GDPR). Guidance on how these regulations interact with open-access data policy[5].
- Type of data, storage, confidentiality, ownership, management of intellectual property and access: procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention, and destruction will be in line with EU standards as described in the Grant Agreement and the Consortium Agreement.

[4] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[5] https://www.openaire.eu/

2.2 BIECO Data Management Strategy

As a project participating in the ORDP in Horizon 2020, the DMP's Data Management strategy of the BIECO project is focused on the observation of FAIR Data Management Protocols. This document addresses for each data set collected, processed, and/or generated in the project the following elements:

Dataset reference and name: Internal project Identifier for the data set to be produced. This will follow the format:

WPNumber-TaskNumber-PartnerName-DataSubset-DatasetName-Version-DateOfStorage

where the project name is **BIECO**, the Partner Name represents the name of the data custodian (WP Lead/ Task Leader).

Dataset description: description of the data generated or collected, including its origin (in cases where data is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the potential for integration and reuse.

Standards and metadata: reference to existing suitable standards. If these do not exist, an outline on how and what metadata will be created.

Data sharing: description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling reuse, and definition of whether access will be open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating the type of repository (institutional, standard repository for the discipline, etc.). In cases where the dataset cannot be shared, the reasons for this will be stated (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Archiving and preservation (including storage and backup): description of the procedures to be put in place for long-term preservation of the data, including an indication of how long the data should be preserved, the approximate end volume, associated costs, and how these are planned to be covered.

2.3 BIECO Type of Data

Among project datasets and deliverables, the following categories of outputs are declared "ORDP" that will be made "Open Access" (to be provided free of charge for public sharing). These will be included in the Open Research Data Pilot and thus be managed according to the present DMP:

- Public deliverables specifically declared as 'ORDP' in the grant agreement such as D1.3 or deliverables that are not Confidential (for example D3.2. Dataset with Software Vulnerabilities is a confidential deliverable which can not be declared as ORDP)
- Articles published in Open Access scientific journal
- Conference and Workshop abstracts/articles

The data collected in BIECO can be split into the following three categories:

- Manually collected data: For example, the data that is collected for verification and validation activities within BIECO, data from the field, such as data regarding

the usage model of a ICT system subject to routine adaptation through the delivering of software updates, and Pictures, audio, and videos (from pilots and workshops).

- Data automatically collected through technology: For example, real-time data of systems sensors and actuators that is collected for developing monitoring tools.
- Personal Information: For example, project partners representatives, or project external individuals who voluntarily participate in the project.

Once generated (or collected), these data will be organized in dataset relating to the category of the data and stored in several formats, which are: a) Documents, b) Images, c) Data, and d) Numerical codes. Considering that DMP is a living document, this version of the BIECO´s DMP does not include a compilation of all the metadata about the data being produced in BIECO project, but there are already several domains considered in the project which follows different rules and recommendations. This is a very early stage identification of standards:

- Microsoft Office 2010 for text based documents such as report, deliverables, publications, and etc. (or any other compatible version): **doc, .docx, .xls, .xlsx, .ppt, .pptx** . Also, especially where larger datasets need to be dealt with, **.csv and .txt** file formats will be used. All finished and approved documents will also be made available as **.pdf** documents.
- Illustrations and graphic design will make use of Microsoft Visio (Format: **.vsd**), Photoshop (Format: different types possible, mostly **.png**), and will be made available as **.jpg, .psd, .tiff** and **.ai** files.
- **PFDs, PIDs** and layouts will preferentially use **inkscape.org**, an open source software for vector graphics. (Format**: .svg**), and will be made available as **.png, .jpg** and **.pdf** files.
- **mp3** or .**wav** for audio files.
- **.avi , .flv, .mov, .MP4,** and **.wmv** for video files.

These file formats have been chosen because they are accepted standards and in widespread use. Files will be converted to open file formats where possible for long-term storage.

Metadata will be comprised of two formats – contextual information about the data in a text based document and ISO 19115 standard [6]metadata in an xml file. These two formats for metadata are chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format).

[6] https://www.iso.org/standard/26020.html

3. FAIR Data

3.1 Making Data Findable, Including Provisions for Metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research. Metadata (the type of data, location, etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

As described in the project Grant Agreement (Article 29.2), the bibliographic metadata includes all of the following:

- the terms “European Union (EU)” and “Horizon 2020”;
- the name of the action, acronym, and grant number;
- the publication date, and length of embargo period if applicable
- a persistent identifier

BIECO open data will be collected in an open online research data repository: ZENODO. Its repository structure, facilities, and management are in compliance with FAIR data principles. ZENODO is an OpenAIRE that allows researchers to deposit both publications and data, providing tools to linking them to these through persistent identifiers and data citations. ZENODO is set up to facilitate the finding, accessing, re-using, and inter-operating of data sets, which are the basic principles that ORD projects must comply with. Zenodo repository is provided by OpenAIRE and hosted by CERN. Zenodo is a catch-all repository that enables researchers, scientists, EU projects, and institutions to:

- Share research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science.
- Display their research results and get credited by making the research results citable and integrating them into existing reporting lines to funding agencies like the European Commission.
- Easily access and reuse shared research results.
- Integrate their research outputs with the OpenAIRE portal.

Search Keywords

Zenodo allows performing simple and advanced search queries on Zenodo using the keywords. Zenodo also provides a user guide with easy-to-understand examples.

Naming Conventions

Files and folders at data repositories will be versioned and structured by using a name convention consisting as follow: **BIECO-Dx.y-YYYYMMDD-Vzz.docx**

Version Numbers

Individual file names will contain version numbers that will be incremented at each revision (Vzz).

3.2 Making Data Openly Accessible

In order to maximize the impact of BIECO research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The BIECO project datasets are first stored and organized in a database by the data owners (personal computer, or on the institutional secure server) and on the project database (project website). All data are made available for verification and re-use unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, the Creative Commons license will be used in some cases. The BIECO dataset deliverables are both public (data access policy unrestricted) and they will be accessible by:

- BIECO project web site [7]
- OpenAIRE
- ZENODO [8] for ORDP data and datasets
- Open access journals

All data deposited on ZENODO are accessible without restriction for public. For other data, potential users must contact the Intellectual Property Right (IPR) team or the data owner in order to gain access. If necessary, appropriate IPR procedures (such as non-disclosure agreements) will be used.

With respect to deliverables of the project, the documents will have different visibility. Some of them have been considered as confidential, and some others will be public. The later will be available on the website in .PDF format. One of the main reasons for the confidentiality of data collected and generated in this project is security considerations.

With respect to algorithms of intelligence and other methodologies. They will be developed with an open source philosophy, and they will be made public when the authors considered it appropriate.

With respect to privacy, the data related contact information of the partners and external individuals who voluntarily participated in the project will not be accessible for public.

3.3 Making Data Interpretable

Partners will observe OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines[9] for CRIS Managers based on CERIF-XML.

[7] http://bieco.org
[8] https://zenodo.org
[9] https://guidelines.openaire.eu/en/latest/

Partners will also ensure that BIECO data observes FAIR data principles under H2020 open-access policy[10].

In order to ensure interoperability, all datasets will use the same standards for data and metadata capture/creation.

As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the DMP. In specific, information on data and metadata vocabularies, standards, or methodology to follow to facilitate interoperability and whether the project uses the standard vocabulary for all data types present to allow interdisciplinary interoperability.

3.4 Increase Data Re-Use (Through Clarifying Licenses)

Creative Common Licensing with being used to protect the ownership of the datasets. Both Share-Alike and NonCommercial-ShareAlike licenses will be considered for the parts of datasets for which the decision of making that part public has been made by the Consortium.

However, an embargo period may be applied if the data (or parts of data) are used in published articles in "Green" open access journals. The recommended maximum embargo period length by European Commission is 6 months.

For datasets deposited on a public data repository (ZENODO) the access is unlimited.

Restrictions on re-use policy are applied for all protected data (see Figure 1: Open access to research data and publication decision diagram), whose re-use will be limited within the project partners.

Other restrictions could include:

- the "embargo" period imposed by journals publication policy (Green Open access);
- some or all of the following restrictions may be applied with Creative Commons licensing of the dataset:
  - Attribution: requires users of the dataset to give appropriate credit, provide a link to the license, and indicate if changes were made.
  - NonCommercial: prohibits the use of the dataset for commercial purposes by others.
  - ShareAlike: requires the others to use the same license as the original on all derivative works based on the original data.

The internal process of Quality evaluation is activated throughout the entire project duration to assess both project data /products and project process (See the D1.2 Project Quality Manual[11]).

An internal peer review is performed for the main project deliverables to guarantee the deliverable is developed with a high level of quality. The process for the submission of the deliverables involves the communication between task and WP Leaders (WPL) and

[10] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-guidelines-oa-covid-19_en.pdf

[11] https://didatec.sharepoint.com/sites/BIECO/Shared%20Documents/General/Deliverables/D1.2%20-%20Project%20Quality%20Manual/BIECO_D1.2_29.12.2020_V1.0.pdf

the project coordinator. The task leader will coordinate all inputs from the other partners involved, chooses at least one internal reviewer [Reviewer(s) from inside the responsible beneficiary for deliverable who was not involved in the creation of the deliverable] and one external reviewer [Reviewer(s) from outside the responsible beneficiary for deliverable], and prepare the draft deliverable. The draft document will be reviewed by the WPL, the internal and external reviewers for the assessment of the contents, technical consistency, and editorial and language aspect and submitted in due time to the Coordinator.

The project data will remain re-usable for at least 1 year after the project.

3.5 DMP Review Process and Data Inventory

The internal process of quality evaluation and reporting is activated throughout the entire project duration to assess both project data /products and project process (See the D1.2 Project Quality Manual). Results data will be also analysed and collected throughout the project's entire duration. To this purpose the Dissemination and Communication Report (See the T9.3: Dissemination, Communication activities, and data management (D9.1, D9.2, and D9.3) will also be filled in by each partner about every six months: it includes the description of articles, papers and scientific publications too. Thus, all research data generated and publications will be analysed and described by using the Data Inventory Table (Annex I), WP leaders and partners authors of publications are required to fill in periodically.

Further updating of the DMP will include the eventually updating of online research data repository where data are collected and shared and the data the description of the dataset and research data gradually generated and collected.

4. Allocation of Resources

Costs related to OA to research data in Horizon 2020 are eligible for reimbursement under the conditions defined in the H2020 Grant Agreement, in particular Article 6 and Article 6.2.D.3, but also other articles relevant for the cost category chosen. Project beneficiaries will be responsible for applying for reimbursement for costs related to making data accessible to others beyond the consortium.

The costs for making data FAIR includes:

- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost-sharing, in the case of multiple authors, shall be decided among the authors on a case-by-case basis.
- Project Website operation: *to be determined.*
- Data archiving at ZENODO and on another online database: *free of charge.*
- Copyright licensing with Creative Commons: *free of charge.*

The project member of the General Assembly is also responsible for the Data Management of the BIECO dataset and research data in accordance with each organization's internal Data Protection Officer (DPO). The General Assembly will be the ultimate decision-making body of the Consortium and responsible for taking major strategic decisions with respect to data management if necessary. It will also promote consensus in case of conflict and, if no consensus can be found, it will take decisions according to the procedures and rules defined in the Consortium Agreement.

Each partner is responsible for the data they produce. Any fee incurred for Open Access through the scientific publication of the data will be the responsibility of the data owner (authors) partner(s).

5. Data Security

Data access and sharing activities will be implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with H2020 rules.

The following guidelines will be followed in order to ensure the security of the data:

- Store data in at least two separate locations to avoid loss of data;
- Encrypt data if it is deemed necessary by the participating researchers;
- Limit the use of USB flash drives.
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.
- Ask users to authenticate for using the data.
- Associate roles to the users to determine the access level.
- Periodic backups

All project deliverables and data will be stored and shared in the Microsoft Team folder restricted to the project consortium. As an initial step, only the Consortium Partners will have access to the cloud storage where dataset and metadata are filed. Following, scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through ZENODO and other databases to promote the data-making FAIR.

6. Ethical Aspects

BIECO partners will assure that the EU standards regarding ethics and data management are fulfilled in compliance with the ethical principles (see Article 34) and confidentiality (see Article 36 as set out in the Grant Agreement). In accordance with the General Data Protection Regulation 2016/679, the data controllers and processors are fully accountable for the data processing operations.

7. Conclusions

One of the main objectives of BIECO's project is the identification of vulnerabilities across complex ICT systems. In this context, the collection and generation of data is key to the identification of such vulnerabilities. This situation makes necessary the development of a DMP not only regarding the research information generated within BIECO, but also regarding the generation of datasets that may contain sensitive information due to security reasons.

The BIECO´s Data Management Plan provided a strategy for managing key data generated and collected during the project and optimize access and re-use of research data. This document describes the main principles and guidelines for the Data Management for the BIECO project. Current document is an incomplete living document which will be updated throughout the project lifetime. Further updating of the DMP will include the eventually updating of online research data repository where data are collected and shared and the data the description of the dataset and research data gradually generated and collected.

8. ANNEX 1- Data Inventory Table

Table 1 Data Inventory Table

Dataset Code/no.	Dataset Name	Open/ Restricted	Data Types	Data purpose	Source of Data	Data creation time	Data finalization time	Responsible person/entity	New/ Existing data	File Formats	Method of Data Capture	Size	Missing Data (%)	Data Utility- Who outside of the project consortium might use the data?	Type of IP/ Protection sought	How will data be re-use	Ethical Issues? Y/N