MarginUp!	**Raising Bio-based Industrial Feedstock of Marginal Lands**	

Data Management Plan

D8.1

Date : 31st May 2023

Author: ATB

Funded by the European Union		Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor REA can be held responsible for them.

Deliverable Information Sheet

Project info

Project Acronym	MarginUp !
Project Title	Raising the bio-based industrial feedstock capacity of Marginal Lands
Grant Number	101082089
Project Coordinator	Philipp Grundmann Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB) Email: pgrundmann@atb-potsdam.de / marginup@atb-potsdam.de
Project Duration	December 2022 – November 2026

Deliverable info

Deliverable No.	D8.1
Dissemination level [1]	PU
Work Package	WP8
Task	8.3
Lead beneficiary	ATB
Contributing beneficiary (ies)	ZIC, ZALF, CluBe, RISE, IFAU, inter3, ATLANTIS
Due date of deliverable	31st May 2023
Actual submission date	31st May 2023

[1] PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)

Document History

V	Date	Beneficiary	Author/Reviewer
1	15.05.2023	ATB	Thi Huyen Trang Dam, Philipp Grundmann
2	19.05.2023	ZIC, ZALF	Rebeca Val, Miguel Maeztu, Michael Glemnitz
3	22.05.2023	ATB, ZIC, ZALF	Rebeca Val, Miguel Maeztu, Michael Glemnitz, Thi Huyen Trang Dam, Philipp Grundmann, James Anderson
4	29.05.2023	ATB, ZIC, ZALF	Rebeca Val, Miguel Maeztu, Michael Glemnitz, Thi Huyen Trang Dam, Philipp Grundmann, James Anderson, Andree Jacobs
5	31.05.2023	ATB, ZIC	Philipp Grundmann, Rebeca Val

MarginUp! in a nutshell

The MarginUp! project proposes solutions to secure use and return profitability on marginal lands while enhancing biodiversity by cultivating climate-resilient and biodiversity-friendly non-food crops for sustainable for industrial feedstock on marginal lands.

Working closely with land managers, farmers, and stakeholders from the growing bioeconomy industry, MarginUp! will create sustainable and circular value chains and increase the resilience of rural farming systems. To further improve biodiversity and environmental benefits, MarginUp! will focus on understanding which marginal lands are suitable with the lowest impact for low indirect land-use change (ILUC) biomass production.

MarginUp! will provide viable outcomes to ecosystems which are water-stressed as a result of climate change, including desertified areas of Mediterranean and Central European member states, as well as contributing to restoring and stimulating ecosystems in abandoned mine lands, and boosting land yield and health in low productivity marginal lands.

MarginUp! is building on learning from seven use-cases: Five implementations across Europe – in Spain, Greece, Sweden, Germany, and Hungary – as well as use-cases in Argentina and South Africa, together increasing the replication potential of the project's results. MarginUp! will identify the best practices for sustainable biomass production and biobased products that safeguard biodiversity and local ecosystems. Each use-case considers the current use and properties of the area and proposes crops and crop rotation strategies that enhance biodiversity and increase soil productivity according to local requirements from Mediterranean soils in Spain to mining lands in Greece, boreal soils in Sweden, wetlands in Germany, desert lands in Hungary, degraded pastures in Argentina, and areas with bush encroachment in South Africa. The proposed crops create a sustainable supply of resources to foster the development of the bioeconomy businesses at local and regional levels while providing ecosystem benefits and building resilience to climate change.

On that basis, the MarginUp! project will enhance European industrial sustainability, competitiveness, and resource independence, by reducing the environmental footprint, including on biodiversity, enabling climate neutrality and increasing resource efficiency (particularly through upcycling and cascading use of biomass) along 5 value chains, and developing innovative bio-based products and enhanced technologies that will lessen EU reliance on fossil-based products.

To stay up to date with MarginUp! project events and reports, follow us on Twitter (@MarginUp_EU), LinkedIn (MarginUp! EU) or visit www.margin-up.eu.

Summary

The Data Management Plan (DMP) describes the collection, generation, management and preservation of data during MarginUp! project, in compliance with Article 15 of the Grant Agreement Number 101082089 which states mandatory the use of open access to scientific publications (Article 15.2), with the exemption shown in Article 15.3.

The DMP is a living document, which will be updated during the project. The first version will be produced in month 6 as deliverable D8.1 and will be updated in month 19 (D8.3) and month 37 (D8.5)

The DMP aims at making data findable, accessible, interoperable and reusable (FAIR) and includes:

- The handling of research data during and after the project;
- The type of data collected, processed and generated by the project;
- The methodology and standards applied;
- Whether data will be shared/made open and how;
- How data will be curated and preserved.

The implementation of guidelines described in this document is the responsibility of all MarginUp! project partners.

Spelling Guidelines

Standardised British Spelling (NOT Oxford Spelling!) should be used in all documents. Generic terms are spelled in lower case, specific terms and proper names are spelled with initial capitals. For metric tonnes use the NOT tons.

Disclaimer

Table of Contents

List of tables

List of Acronyms

CA	Consortium Agreement
CERN	European Organization for Nuclear Research
DMP	Data Management Plan
DP	Privacy Policies
EC	European Commision
EU	European
GA	Grant Agreement
GDPR	General Data Protection Regulation
GPG	Gnu Privacy Guard
GRACE	Name of a EU Horizon project
HE	Horizon Europe
ILUC	Indirect Land Use Change
MAGIC	Name of a EU Horizon project
MIDAS	Name of a EU Horizon project
ORDP	Open Research Data Pilot

Keywords List

- MarginUp!
- Data management plan
- Fair data
- Data security
- Data quality

1. Introduction

Projects under the EU's Horizon European (HE) program must specify what kind of information and results will be generated or collected during the life of the project and when and how they will be stored and published. The main objective of the Data Management Plan (DMP) presented in this deliverable report D8.1 is to provide guidelines for including the main elements of the data management policy that are going to be applied by the MarginUp! project consortium. The DMP guidelines of the MarginUp! project will ensure research data is Findable, Accessible, Interoperable and Reusable (FAIR data principle) to make possible that knowledge is integrated and available for re-use in future research and projects. The Grant Agreement (GA) for the MarginUp! project provides for a first version of the DMP by M6 of the project (Deliverable D8.1) with a definitive update at the end of the project.

The DMP has been structured by following the Guidelines on FAIR Data Management in HE provided by the European Commission in May 2021. The DMP describes the types of data that will be generated and/or gathered during the project, the standards that will be used, the ways in which data will be exploited and shared (for verification or reuse), and in which way data will be preserved. This DMP has been prepared by taking into account the template of the "Data Management Plan Template HE"[1]. The elaboration of the DMP will allow MarginUp! partners to address all issues related with data protection, including ethical concerns and security protection strategy. MarginUp! takes part in the Open Research Data Pilot (ORDP); this pilot aims to improve and maximise access to and re-use of research data generated by HE projects, such as the data generated by the MarginUp! platform during its deployment and validation. Moreover, under HE each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results: these publications shall be made also available through the public section of the MarginUp! website. All these aspects have been taken into account in the elaboration of the DMP.

As stated in "A European strategy on the data value chain", the intelligent use of data enables the creation of new products and services, and has the potential to transform Europe's service industries and significantly increase their efficiency. In the public sector, it will lead to cost reduction of operations, increase of efficiency and better and more personalised services for citizens. Thus, the aim of MarginUp!

[1] https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fec.europa.eu%2Finfo%2Ffunding-tenders%2Fopportunities%2Fdocs%2F2021-2027%2Fhorizon%2Ftemp-form%2Freport%2Fdata-management-plan_he_en.docx

DMP is to provide an analysis of the main elements of the data management policy that are going to be used by the consortium.

The DMP will be a living document throughout the project, and this initial version will evolve during the MarginUp! lifespan according to the progress of project activities. This report describes the procedures of data collection, storing and processing, with a final overview on MarginUp! security protection strategy. This report does not cover the general concerns related to ethics and data protection, as they are the focus of dedicated deliverables.

The main responsible partner for the development of the DMP is ATB. This first version of the DMP presented in this report was prepared in collaboration with the following partners from the DMP-group including ZIC, ZALF, RISE, ATLANTIS, IFAU and CluBe. However, the implementation of the DMP is the responsibility of all project partners, especially those involved in the collection, storage, organisation, maintenance, use and publication of data, information and results within the MarginUp! project.

2. Data summary

2.1. Purpose of data collection/generation

The main objective of MarginUp! is to introduce climate-resilient and biodiversity-friendly non-food crops for sustainable industrial feedstock in marginalised and low-productivity land, resulting in robust and sustainable value chains that benefit both the local biodiversity and ecosystem services as well as its regional socio-economic positioning, by increasing the farming system resilience in rural areas. In this perspective, MarginUp! will generate technical data and collect technical, commercial, ~~persona~~social, economic and environmental data at all levels from individuals and organizations to networks and societies in compliance with all national and EU ethics and legal requirements in the frame of the following activities:

- analyze the availability of marginal lands, the potential of biobased products and farming systems to be implemented;
- engagement of end-users and stakeholders along the value chains;
- support the use cases in finding the right balance between productivity and ecosystem services, especially about biodiversity and social sustainability goals;
- implement sound agronomic practices in marginal lands and compare it against current practices. 5 demonstrators in different EU countries: Germany, Sweden, Spain, Hungary, Greece and 2 associated demonstrators in South Africa and Argentina;
- best combination of technical solutions and practices for the new industrial value chains including biotechnology approaches, and better application of cascading use of different feedstocks;
- engage and understand the different relevant stakeholders and all actors in the implemented use-cases and obtain experience exchange, replication, and mutual learning, both regionally and internationally.

Only data that is needed to perform project activities will be collected, and as far as possible, participants will not be asked to provide personal data unless this is necessary.

2.2. Data identification and description

2.2.1. Types and formats of data

Data will be organised in datasets relating to the category of the data and site of collection. Dataset units will be collected, analysed and produced during the project life span. Each data set will contain, at least the work package to which it belongs and data typology, with a short description. MarginUp! envisages the collection of external data to be analysed. The external datasets and the datasets resulting from personal surveys and interviews will be stored ensuring the confidentiality of participants' data. By default, in compliance with GDPR (EU GDPR 2016/679) all project data will be handled with rigorous care.

MarginUp! will compile and produce project data sets of long-term value in the context of regional implementations and assessments. A dataset can include different types of formats. Types and formats of MarginUp! data are presented below:

(i) Data from use cases include:

- EU marginal lands regards, habitats, biodiversity, land use types and farming systems and best management practices. In addition, involvement of the engagement of number of actors/stakeholders in order to create sustainable and replicable business cases (e.g. risks, barriers, social acceptance);
- Feedstock and biomass performance data (e.g. capacities, productivities, technical conditions), estimated total cost of production and gross margin.
- Product data (e.g. product descriptions, compositional data, quality data, market price); i.e. description of the use cases and the related value chain and its partners, including the number of the involved end-users;
- Business objectives: motivation and reasons for the use cases exploitation; limitations and risks of the use cases; scope and impact of the use cases for a region or other field of application; Customer and user needs, value proposition and market data; service or product characteristics, management team profile;
- Local stakeholder data and information at the use case sites: publicly available data necessary for mapping and engaging stakeholders and stakeholder groups, data collected in Stakeholder Boards (pending GDPR consent) on needs with regard to functionality and design or solutions and opportunities for implementation of novel technologies, workshop evaluation data and data obtained from surveys and personal interviews.

Collected data from use case will be qualitative and quantitative data, with text formats (PDF/A, .txt, .doc/.docx), Spreadsheets (.xls/.xslx), Pictures (jpg, png), Shapefile (.shp), Databases (.cvs; accdb), AutoCAD (.dwg), WEBM File and other.

(ii) *The analysis/assessments output data from WPs/tasks* will be qualitative and quantitative data with different formats. Output data includes: data and information generated from: 1) Mapping marginal lands for low ILUC biomass production and materiality analysis to identify the relevance to stakeholders (WP1); 2) Biodiversity assessment (WP2); 3) Analysis of the specific components for a successful implementation of the 5 use cases (WP3); 4) Environmental, social and economic impact assessments (WP4); 5) Stakeholders analysis (WP5); and 6) Economic analyses of the industrial feedstock value chains and roadmap for scaling up value chains (WP6).

The type and format of the output data shown in Table 1 relate to successive and interrelated WPs/tasks. The output data formats will be coordinated between the work-packages and tasks to ensure the use of output from one WP/task as input for another WP/task. Examples for data and information outputs shared between tasks and work packages as shown in the table below are data from (1i) that is used as input in (1ii), (5) and (6). Further, (2) delivers output data to (4) and (6), and (3) uses input from (1i). The results of assessments and analysis of (1i), (2), (3), and (4) will contribute to (6).

Table 1.Output data and formats

Output data	Formats	Collected/Generated data sources
(1i) Mapping marginal Lands for low ILUC biomass production, (1ii) Materiality analysis	.PDF/A, .txt, .doc/.docx, .xls/.xslx, JPEG, GIF, HTML, WEBM File, GIS (shp, bdf, shx, SQL) and other.	Collected data from use cases by questionnaires, interviews, workshops; generated output data from GIS soft wares and materiality assessment software.
(2i) Biodiversity assessment	.PDF/A, .txt, .doc/.docx, .xls/.xslx GIS (shp, bdf, shx, SQL), metadata and other.	Collected data from use cases by questionnaires, interviews; generated data from GIS soft wares.
(3i) Analysis of the specific components for a successful implementation of the 5 use cases	.PDF/A,. txt, .doc/.docx, .xls/.xslx, photos (JPEG, GIF), GIS (shp, bdf, shx, SQL) and other.	Generated data from use cases questionnaires, interviews, workshops.
(4i) Environmental and Social economic impact assessments	.PDF/A, .txt, .doc/.docx, .xls/.xslx and other.	Generated output data from calculate models VCA (Value chain analysis), CBA (Cost Benefit Analysis), sLCA (Social Life Cycle Assessment).
(5i) Stakeholders analysis	.PDF/A, .txt, .doc/.docx, .xls/.xslx and other.	Collected/Generated data from questionnaires, interviews, workshops.
(6i) Analyse the economics of the industrial feedstock value chains and road map for scaling up value chains	PDF/A, .txt, .doc/.docx, .xls/.xslx and other.	Collected/Generated data from questionnaires, interviews, workshops, group discussion.

(The table will be updated in the next versions of DMP)

(iii) Data collected/generated related to Communication, Dissemination & Replication Activities (e.g. From the MarginUp! Network, MarginUp! website (e.g. cookies) …), with text formats (PDF/A, .txt, .doc/.docx), formats of Pictures (jpg, png), Spreadsheets (.xls/.xslx), Video (avi, flv, mov, mp4, wmv).

(iv) Periodic documents/Reports/Publications will be generated in MarginUp!, with text formats (.PDF/A, .txt, .doc/.docx).

(v) Contact Data from MarginUp! partners: names, titles, job levels, departments, LinkedIn profiles, email addresses, and phone numbers.

(vi) *Other data related to project management (WP8) (e.g.* Conferences and Workshops abstracts) has .doc, .docx, .xls, .xlsx, .pdf, .ppt, .pptx,.jpeg formats.

2.2.2. Size of data and expected volume

Size of Data

The expected size of individual datasets largely ranges between a few MB to GB, with a typical value of about 150 MB per file. Dataset sizes will continue to be evaluated during the project and will depend on the extent and the nature of the data that are made available.

Expected volume

The expected data volume will be approximately 100 GB. The ATB storage which is project overhead will cover this size with no additional costs. Exchange of data with project members at different sites will be secure and redundant through use of available ATB cloud storage**.**

2.2.3. Data Utility

The data collected by MarginUp! will be suitable for use by:

- MarginUp! consortium
- Stakeholders involved in the fields of the project
- Scientific community
- European Commission services and European Agencies
- EU national bodies
- General audience

2.2.4. Origin of data

MarginUp! will collect data at 5 main locations - EU use cases: Spain, Hungary, Greece, Sweden and Germany and 2 international use cases: South Africa and Argentina . Depending on the type of data, there will be various methods and origins of data collection involved at each site.

For manually collected data the main origins will be:

- National/regional land use and biodiversity strategies/plannings
- Soil, water and habitat maps
- Interviews with groups and individual participants at each site
- Feedback from participants at stakeholder workshops
- Survey responses
- Literature study/review and open data (re-use of existing data)
- Personal data upon requests of the consortium

2.2.5. Re-use of Existing Data

Data will be sourced from several already existing datasets and/or reports and will be updated making use of the knowledge of project partners.

The re-use of existing data available from research projects and from other European projects is being encouraged. References from Horizon projects such as data inheritance from MAGIC project 'Marginal land for growing industrial crops' (2017-2021), GRACE project 'Growing advanced industrial crops on marginal lands for biorefineries' or data exchange from ongoing MIDAS project 'Marginal lands, industrial crops and innovative bio-based value chains' (2022-2026) would be considered as advantageous re-use data source. This section will be updated at a later stage of the project.

3. Data manager(s)

MarginUp! DMP will be stored as a project deliverable document, with regular updates and will have successive versions. The first version of the project document MarginUp! DMP will be completed end of May 2023 by integrating the discussed procedures of the work-packages and their interfaces. The responsible team members for producing the MarginUp! DMP and for coordinating the data management is ATB, all work-packages contribute to the MarginUp! DMP and support the data management by collecting project relevant data, providing data for the use in other work-packages and deliver information about the data sets for recording in the MarginUp! DMP. Each work package may have more than one data manager depending on number of different tasks‘ manager, which require data managing efforts. Data manager(s) of work packages’ tasks identified are as below:

Work package 1 (WP1): RISE, ATB, Inter3, ZALF

Work package 2: ZALF

Work package 3: CluBE, RISE, ATB

Work package 4: ATLANTIS, ATB

Work package 5: Inter3, ATB

Work package 6: IFAU, INNOMINE, ATLANTIS

Work package 7: Greenovate!, REVOLVE

Work package 8: ATB, ZABALA

4. Data organization and exchange (Within Consortium, during project)

During the project the developing, implementing, overseeing and updating of MarginUp! project DMP is integrated in project management (WP8, Task 8.3). The responsible team member (see section 3) receives contributions from all work-packages regularly. During the project, MarginUp! will use ATB central unit's IT infrastructure for exchange between the work-packages and to protect value of long-term data.

A Nextcloud-repository for MarginUp!, located at the server at ATB, has been set up to handle effectively the documentation of the project. This tool is aimed at working on collaborative documents and sharing final documents of common interest.

This will be the project's online collaboration platform during the project lifetime. It provides a user-friendly and safe access to data through a web interface and a platform to view, sync and share the files across devices easily — all under user's control. It uses an open architecture, extensible via API for applications and plugins and works with any storage.

Nextclould file is accessible only for MarginUp! work-packages members, included daily backups, high standards of data security. All partners of the consortium will have access to the project Nextcloud file and able to up- and download datasets. The availability of datasets, which will be delivered from one work-package as input for another work-package will be organized by the project management and is pre-defined in the project time-schedule. Responsible for data consistency, quality and storage are the partners of the project consortium during the project. To organize the exchange and ensure the storage of project data, especially of long-term data, the uploaded data on Nextcloud will be backup at ATB's IT-infrastructure resources.

In Nextclould repository, the metadata of datasets are made publicly accessible for all partners of the consortium. This aims at better transparency for MarginUp! research as well as attracting more partners attention, cross checking between members and possible further research collaboration. Per the policies outlined in Section 7.2 of MarginUp's data security protocol (including deployment of a threefold security protection strategy; authentication, authorization, and encryption; focusing on data aggregation and pseudonymisation techniques; preventing Internal threats and human errors), any observation data or primary data generated or collected by members of MarginUp! must be made accessible to other members during the project period. Any members interested in accessing this data must also abide by MarginUp!'s data security policy. However, certain partners may request to withhold generated data for

a limited period of time, specifically for publishing purposes. In such cases, each data controller is required to provide a detailed explanation regarding the purpose and length of time for which the data will be withheld.

5. Fair data management

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. All project deliverables and data will be stored and shared in the team folders restricted to the project consortium. As an initial step, only the consortium partners will have access to the project cloud storage where dataset and metadata are filed. Following, scientific publications and articles, the dataset deliverables and the final research results will be stored and published through Zenodo and other repositories (e.g. OpenAgrar) following the FAIR principles.

5.1. Making data findable (/including provisions for metadata)

5.1.1. Data set reference, names and versions

Each file is assigned with a unique name which helps to distinguish and easily identify datasets. This name can also be used as the identifier of the datasets. We use the following practice in order to create the dataset names. Each designed dataset name consists of three different parts separated with a ”.” (dot) character:

ProjectName.DatasetName.Version, where

1. The DatasetName represents the full name of the dataset.

2. The ProjectName is MarginUp, in order to clearly identify for all datasets the origin (The exclamation mark (!) will be removed because of the known problems that occur if this special character (!) in Access).

3. The Version of the dataset represents in which phase of the project the dataset was released:

i. DB the live database during project lifetime.
ii. RP1Export export of the database at M15.
iii. RP2Export export of the database at M27.
iv. FinalExport export of the database at the end of the project, i.e. M48.

An example of a dataset’s name could be the following: AreasMarginalLand.MarginUp.RP1Export.

In order to catalogue data in the repository as well as to facilitate their search and re-use, metadata will be filled in when uploading datasets in Zenodo[2] (or OpenAgrar[3])(see section 5.1.4 for further information) encompassing at least the following elements:

- Title
- Author/s
- Publication Date
- Description
- Access right
- License
- Funding

Additionally, all open data, publications and open source software deposited in the Zenodo repository (see 5.1.4) will use DOI versioning. DOI versioning allows for updating a dataset after it has been published and to cite either a specific version of a dataset or all versions of a dataset (see https://blogs.OpenAIRE.eu/?p=2010).

5.1.2. Naming conventions

File names will include at least a version number and/or a time stamp. Files and folders at data repositories will be versioned and structured by using a name convention consisting as follow: YYYYMMDD MarginUp. DatasetName.Vzz.docx.

5.1.3. Standards and metadata

Providing as much accurate and rich metadata as possible will improve the data's findability and re-usability. Metadata for describing the data that is collected and generated by the MarginUp! project is needed for facilitating open access to the data.

MarginUp! records metadata for each produced data set. All MarginUp! datasets will use a standard format for metadata. Each dataset description will specify the Data and metadata standards used. In this regard, the governance of metadata is a relevant part of the Data Management Plan. This is because even metadata that contains no obvious identifiers could qualify as personal data, as rich metadata could lead to reidentify a data subject when combined with other data sources.

[2] https://zenodo.org/
[3] www.OpenAgrar.de

Several metadata standards exist (see http://www.dcc.ac.uk/resources/metadata-standards/list or https://rdamsc.bath.ac.uk/) and it may not be possible to find one that fits all purposes. Therefore, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories and to use data-type specific schema extensions, if necessary.

The Zenodo deposition metadata domain model (http://developers.Zenodo.org/#representation) which is based on DataCite's metadata schema (https://schema.datacite.org/) minimum and recommended terms will be used for open data generated by the project and deposited in an appropriate repository (see 4.2.5). We will use these standards for the release of data outside of the project.

For MarginUp!, the following deposition metadata fields are mandatory:

• title (of the deposition)

• description (abstract or description for deposition)

• files (Deposition files identifiers, filenames, size of the files in bytes and MD5 checksum of files)

• upload_type (Type of the deposition from a controlled vocabulary (publication, dataset, software, …)).

• publication_date (Date of publication in ISO8601 format (YYYY-MM-DD)).

• creators (The creators/authors of the deposition)

• license (Open license from controlled vocabulary “Open Definition Licenses Service”).

• doi (Digital Object Identifier assigned by the DOI registrant (e.g. Zenodo), also used for versioning (see 2.1.4)

• keywords (descriptive to the content of the dataset)

• related_identifiers (Persistent identifiers of related publications, datasets and software (see 2.2.6)).

• communities (List of communities the deposition to appears in (https://zenodo.org/communities/clarity/))

• grants (List of European Commission FP7 grants which have funded the research for this deposition (730355). Needed to establish the relationship to MarginUp! in OpenAIRE)

Trying to improve metadata and the obtained results, MarginUp! participant will consider mentioning controlled vocabulary that significantly enhances metadata and increases the findability of the data. For many disciplines, there are already specialized classifications and thesauri (see www.bartoc.org) as well as standard data for persons, institutions or research funding agencies. A set of general keywords that should apply to all public datasets, scientific publications and public deliverables will be defined and updated to next version and DMP.

5.1.4. 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. 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. A data inventory is a centralized metadata collection, indicating all of the datasets the project collects and maintains.

The Data Controllers at each pilot site will be responsible for managing the completion of the metadata, uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must be descriptive to the content of the dataset.

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..

5.1.5. Archiving, Preservation and Data quality assurance

Archiving and preservation

The MarginUp! partners agreed on the procedures that will be used in order to ensure long-term preservation of the datasets.

Datasets will be stored on Zenodo (https://zenodo.org/) a generic repository for EC funded research developed by CERN and launched in May 2013. To be an effective catch-all, that eliminates barriers to adopting data sharing practices, Zenodo does not impose any requirements on format, size (currently accepts up to 50GB per dataset), access restrictions or license.

In addition, datasets stored on Zenodo are automatically part of OpenAIRE (https://www.openaire.eu/), the EC-funded initiative which aims to support the Open Access policy of the European Commission via a

technical infrastructure, thus integrating them into existing reporting lines to funding agencies like the European Commission.

Archiving on Zenodo is free, thus eliminating costs. In relation to this, Zenodo's Terms and Conditions establish that the uploader is responsible for making sure the content is suitable for open dissemination, namely, that it complies with all applicable regulation, including data protection, privacy and intellectual property laws. All relevant stakeholders in the MarginUp! programme will be warned of this requirement and requested to act accordingly. In particular, and without prejudice of additional measures, this may require a process of anonymization of the uploaded contents of survey respondents.

In addition to Zenodo, the generated information can be also uploaded in a more topic related repository like OpenAgrar. OpenAgrar publication server and features research data is centralised in Germany. ATB, as project leader is considering acquire an access to this repository to include the research data linked to subjects as: Plant Sciences, Economy, Biology, Life Sciences, Agriculture, Forestry, Horticulture and Veterinary Medicine.

Data quality assurance

MarginUp! is committed to deliver quality data and adopts data quality assurance procedures to achieve this goal. Quality control of each dataset is the responsibility of the relevant WP leader, supported by the Project Coordinator. Depending on the case, "quality" might have different meanings, depending on the utility and on the re-usage scenarios of the dataset. Data quality assurance might hence imply editing and moderation, cleaning, pre-processing, adding metadata, transforming to a more convenient format or providing easier access. Information about the consortium's efforts to address data quality issues is hence provided for each type of dataset.

In this regard, all stakeholders acknowledge and agree to comply with the traditional principles relating to data quality, which have been updated by the GDPR as principles relating to the processing of personal data, and include, among others, that data shall be processed lawfully, fairly and in a transparent manner, collected for specified purposes and adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Lastly, data quality shall also refer to the dataset as a whole, ensuring population subgroups will not be discriminated or otherwise harmed due to a misrepresentation in the dataset.

5.1.6. Publications and key words

Publications

The publications issued during the project include the Grant Number, acronym and a reference to the Horizon Europe Programme funding, including the following sentence:

"**This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement N° 10108208".**

When displayed together with another logo, the EU logo will have appropriate prominence.

Search keywords

When uploading the data and publications in Zenodo, data owners will enter key words in order to facilitate the search. All open MarginUp! results deposited in a repository will provide search keywords together with their metadata (see 5.1.3). Keywords for open data will be selected from controlled vocabularies that are suitable for the specific type of the data (see 5.3). The list of key words will be compiled during the project and included in the second versions of the DMP.

Keywords used so far: EU Marginal lands; low ILUC value chains; biobased products; biodiversity opportunities; Farming systems resilience; industrial feedstock; farming system resilience; environmental and socioeconomic impact assessment; innovation strategy; biomass production; life cycle assessment; optimisation of biomass products.

5.1.7. Digital Object Identifiers

All open data, publications and open source software produced in MarginUp! (open MarginUp! results) will identifiable and locatable by means of a persistent Uniform Resource Locator (URL). If possible, open MarginUp! results will be assigned a Digital Object Identifier (DOI) in order to make content easily and uniquely citable. Thereby, MarginUp! relies on external services, since DOIs can only be assigned by DOI registrants through a DOI registration agency (see https://www.doi.org/doi_handbook/8_Registration_Agencies.html).

Open MarginUp! results that are deposited in the MarginUp! default Open Access repository (Zenodo, see 5.1.4) will be assigned a DOI automatically and will benefit also from Zenodo's DOI versioning support.

Open MarginUp! results that are deposited in institutional repositories, repositories of scientific publishers or other data and research repositories will be at least indefinable by a persistent URI. If the

institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will be assigned, too.

Whether scientific publications will be assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc. depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository. Zenodo (http://help.zenodo.org/features/) is for example, one of the open data repository repositories that can generate DOIs for research results.

5.2. Making data accessible

5.2.1. Open Access

The H2020 Open Access Mandate aims to make research data generated by HE projects accessible with as few restrictions as possible, but also accept protection of personal or sensitive data due to privacy concerns and/or commercial or security reasons. Open Access in Research and Development projects typically focuses on access to "scientific information", which refers to two main categories:

- Peer-reviewed scientific research articles (published in academic journals);
- Scientific research data (data underlying publications and/or raw data).

In line with H2020 Guidelines on Open Access to Scientific Publications[4], open Access will be implemented in peer-review publications (scientific research articles published in academic journals), conference proceedings and workshop presentations carried out during and after the end of the project. Published articles (accepted author version) will be made public on the MarginUp! website.

Two main routes exist for open access to scientific peer-reviewed publications[5]:

Self-archiving (also called "Green" Open Access) means that the published article or the final peer-reviewed manuscript is archived by the researcher – or a representative - in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed ("embargo period") as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period. Depending on the journal selected, the publisher may require and embargo period between 6 and 12 months.

[4] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[5] European Commission. IPR Helpdesk. Fact Sheet Open Access to scientific publications and research data in Horizon 2020

Open Access publishing (also called "Gold" Open Access) means that an article is immediately provided in Open Access mode by the scientific publisher. The associated costs are shifted away from readers, and instead charged to (for example) the university or research institute to which the researcher is affiliated, or to the funding institutions supporting the research. When using this model, the costs of publishing are not assumed by readers and are paid by the authors, this means that these costs will be borne by the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research. Gold Access will be opted for where possible within the MarginUp! project.

Open Access will not affect the intellectual property generated by research results, as the decision on whether to publish Open Access documents will follow the procedure as described in article 16 of the Consortium Agreement to first seek protection for intellectual property rights [6] as illustrated in the figure below:

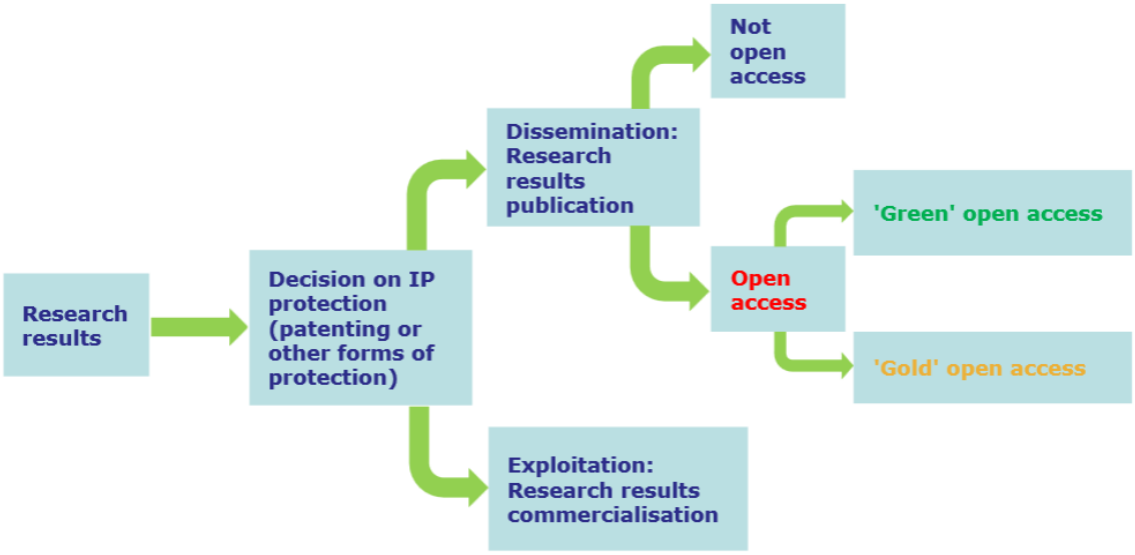

Figure 1: Open access rule for scientific publication and research data in the wider context of dissemination and exploitation

[6] European Commission. IPR Helpdesk. Fact Sheet Open Access to scientific publications and research data in Horizon 2020

5.2.2. Overview of access to data

Table 2. Overview of access to data

Open access		Confidential access	

Data type	**Access**	**Main exploitation guidelines**
Data from the use cases and management practices		Open to public in final reports, project website and in press releases.
		Confidential access for contractual terms, financial Information and management plans.
		Confidential access for personal information in compliance with GDPR.
End-user/personal data[7]		User data will be treated with confidentiality. Disclosure is conditioned by the systematic authorization of end users provided in the "informed consent form" and "information sheets". The collection, storage, protection, retention will be done in compliance with EU Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [8]and in compliance with the 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

[7] For complete Information on Protection of Personal Data, please refer to D7.1 "Protection of Personal Data" - Ethics requirement n°1.

[8] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&from=FR

		Regulation)[9].
Research data from use case sites and the final Road map.		Open access. 3 steps procedure: 1. assessment of the IPR protection 2. review of Inclusion of personal data 3. upload In Zenodo repository
Data collected for the Databases		Open to public in final reports, project website and in press releases.
		Confidential access for contractual terms, Information and business plans.
		Confidential access for personal information in compliance with GDPR.
Data exchanged among partners		contact details (see 2.2.1), File Transfer Protocol (FTP) and secured collaborative space (Gdrive).
Conference presentations, workshops, events, deliverables		Open access in the project website
Peer-reviewed publications		Open access in project website, Zenodo repository and in partners' own repositories[10]

Accessible datasets will be licensed with a Creative Commons (2023) License (https://creativecommons.org/). We plan parallel open access data publication and open access scientific papers during project lifetime and after the project. We intend to publish hybrid open access in journals which allow immediate open access to articles on condition that an 'Article Processing Charge' is paid or golden open access in Open Access journals.

[9] 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), https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504

5.3. Making data interoperable

This section presently can only be partially answered, as the data collection is still in progress. However, most datasets collected from use cases are distributed in the ATB Nextcloud. The Nextcloud system is an open source system that is able to work with a large variety of computing devices and operating systems, with facilities that enable easy access to read or upload datasets. The section will be compiled during the project, as datasets will be made available for MarginUp! and shall include information on:

- Use of open software applications
- Use of standard vocabularies and methodologies

It will also recommend the use of non-proprietary, open formats, nothing vendor/software specific (e.g. AutoCAD, SPSS, MaxQDA). Even when using "common" software packages such as Microsoft Office, it is not possible to trust for backward compatibility and long-term readability. For this reason, files should be saved in open formats at least additionally, if possible. The files to be archived should be unencrypted, uncompressed, patent-free and created in an open, documented standard. This is also in line with FAIR Data's approach to interoperability and re-use of data.

Also see https://en.wikipedia.org/wiki/List_of_open_formats

Zenodo uses JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition (http://opendefinition.org/)), funders (FundRef(f (https://www.crossref.org/services/funder-registry/)) and grants (OpenAIRE (http://api.OpenAIRE.eu/)). Reference to any external metadata is done with a resolvable URL.

5.4. Reusable data (through clarifying licenses)

According to the European Commission, “research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion, or calculation”. Open access to research data is the right to access and reuse digital research data under the terms and conditions set out in Article 16 of the Grant Agreement.

This section will be compiled during the project as datasets will be made available for MarginUp! and shall include information on:

- Licensing of data

- Availability of data and embargo period
- Re-use of data by third parties
- Data quality Insurance processes
- Duration of data for re-use

Open Data and Open Source Software

Data and software are owned by the beneficiary that generates them. Notwithstanding the above owners of open results arising from the MarginUp! project are encouraged to release their work under a Creative Commons license, preferably Creative Commons Attribution 4.0 (CC-BY-4.0, http://opendefinition.org/licenses/cc-by/).

Scientific Publications

Authors of scientific publications arising from the MarginUp! project are encouraged to seek an agreement with the scientific publisher of the publication (see 5.2.1) that allows the authors to

• retain the ownership of the copyright for their work and to

• deposit the publication in an Open Access repository

with Creative Common Licensing being used to protect the ownership of the datasets. 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 rule for scientific publication and research data in the wider context of dissemination and exploitation and see Table 2: Overview of access to data), 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.

6. Allocation of resources

6.1. Costs

There is no fee to use the ATB infrastructure of the central service. The project budget covers (i) maintaining depository, (ii) personal cost, and (iii) for developing, implementing, overseeing and updating the DMP.

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.

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).

Any unforeseen costs related to open access to research data in HE are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement, in particular, in Article 6 and Article 6.2.D.3. Additional details will be reported, as needed, in future versions of the DMP.

6.2. Data management responsibilities

ATB will be manager of the data repository in Nextcloud and their responsibilities shall include:

- Definition, creation, update of repository structure;
- Co-creation with the project partners of the data repository's folders/sub-folders for each user group and document type (e.g. data, metadata, templates);
- Timely upload of data / publications in the repository following validation by the PMB;
- Ensure appropriate versioning, metadata, access and level of dissemination;

- Provide final approval to upload content to the repository.

All datasets described by the project are shared among the Consortium partners of the proposal under a Nextcloud structure. In addition, read only access will be provide to the EC and Reviewers via direct link. The coordinator makes sure to regularly update the access list of personnel within the teams of the partners to the files and sensitive information contained on them.

Data management activities concern the whole project and needs to be coordinated and monitored both at project and work package level. Data management is also linked to publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities can be identified: ATB as the Project Data Manager (WP8 leader) is responsible for:

• developing the data management plan and policy in cooperation with coordination team and data management team and ATB IT department;

• coordinating the technical realisation in WPs (data survey, data repositories, metadata catalogues ...);

• monitoring data management activities (both collection and publication) and deadlines and sending reminders to WP data managers;

• providing support to WP data managers;

• creating the data management plan (D8.2);

• providing solutions for specific issues in accordance with project management;

• coordinate with the Dissemination Manager to responsible for offering assistance in choosing the right publication path (green or gold open access), offering customized help and further guidance for publishing scientific publications, ensuring that the open access policy of the journal complies with the H2020 open data requirements before the researcher submits a manuscript, monitoring that green access (self-archiving) publications are deposited in repositories and sending reminders to partners, monitoring that research data related to a publication is made available in repositories and linked to respective publication, monitoring possible embargo periods and sending reminders to partners, monitoring that publications available.

The Workpackage Data Managers (see section 3) are responsible for:

• the implementation of the data management policy in their respective WPs/tasks;

• monitoring data management activities and deadlines and sending reminders to partners;

• offering customized help and further guidance for filling out the WP data surveys;

• asking partners for missing information or clarifications;

• providing input to the data management plan by analysing and summarising the WP-specific data surveys;

• offering customized help and further guidance for publishing open data and open source software;

• monitoring that open results (data and software) are deposited in the default repository sending reminders to partners;

• contacting the project manager and ethics advisor in case of questions and ethical and privacy issues that may forbid a publication of the data.

7. Data security

7.1. Data storage and back-up

ATB Nextcloud service is administered by the ATB IT group, which follows the data security standards of the German Federal Office for Information Security, including maintaining regular backups of data, with copies stored in multiple locations. To ensure the storage of project data, especially of long-term data, the uploaded data on Nextcloud will be back-up at ATB IT-infrastructure resources. The ATB' resources have daily weekly, and monthly back-up schedules and IT group at ATB, which has experiences in data issues, will responsible for data storage and back-up. Team ATB IT member Christian Schmiel will be responsible for backup and storage of MarginUp! in ATB-infrastructure.

Individual datasets will be simultaneously stored in partners' local IT- infrastructure and updating information of local storages will be performed by responsible manager(s) in each WP/task.

7.2. Data security within Consortium

7.2.1. Deployment of a threefold security protection strategy

The process of collecting, processing, storing data might hide some pitfalls. To reduce the risk of potential malevolent, criminal and/or terrorist abuse, which might be perpetrated also by malicious people authorized to access the information, the MarginUp! Consortium is examining the deployment of a threefold security protection strategy regularly during project life span:

1. by ensuring that the employed security layers and privacy-preserving measures will work properly, keeping access logs and following best practices for system administration;

2. by employing techniques to prevent information leakage "on-the-fly", i.e., through the adoption of the anonymization and pseudonymization approach of personal and sensitive information at collection, communication, and storage time (e.g. via an encryption scheme, hash functions, and/or tokenization). Such an approach will neutralise eavesdropping and/or similarly dangerous hack attempts in the unlikely event of successful retrieval, since it will secure data, making them completely meaningless to the possible attacker.

3. by employing good practices in case of the unlikely scenario of data breach, regarding notification to the compelled data protection authorities and proceeding according to the General Data Protection Regulation.

7.2.2. Authentication, authorization, and encryption

According to the General Data Protection Regulation (GDPR) the implementation of both computerized authentication and procedures for managing authorization's credentials is required. To assure the security of and the trust in the system, and to properly protect the rights and freedoms of natural persons, it is fundamental to provide technical solutions to ensure data security regarding the services offered by MarginUp!. For identity management and data protection mechanisms, MarginUp! will follow the standard practice in the security research community.

Identity management deals with identifying individuals (authentication) and controlling access (authorization) to resources in a system. All the Privacy Enhancing Technologies associated with identity management aim at identity verification with minimum identity disclosure, and protection against identity theft. Due to internetworked services and in general to Cloud technology, the need of a secure identities management has grown increasingly. Identity and access management (IAM) is the security and business discipline that "enables the right individuals to access the right resources at the right times and for the right reasons". It addresses the need to ensure appropriate access to resources across increasingly heterogeneous technology environments and to meet increasingly rigorous compliance requirements. Technologies, services and terms related to identity management will be analysed by the consortium and applied, if applicable for the external services and platforms used during project implementation.

More specifically, following the "Privacy by default and by design" principles, included in the GDPR, the MarginUp! website will adopt an integrated and multilevel approach to protect the user information from the fraudulent access and consumption.

For ensuring the privacy and security of datasets generated by MarginUp! project that will remain private, the GPG (GNU Privacy Guard) will be used. GPG allows encrypting files using public-key cryptography. GPG uses user's private key for signing and encrypting files that can be later decrypted using user's public-key. Anyhow, in MarginUp! project GPG's symmetric encryption algorithm will be used, which allows encrypting files using a password known by all consortium members allowed to check those files. In addition, those encrypted files will be stored at NextCloud, granting access only to proper consortium members.

7.2.3. Focus on data aggregation and pseudonymization techniques

Personal and sensitive data will be made publicly available only after an informed consent has been collected and suitable anonymization techniques have been applied. Before starting the project activities, a careful investigation on privacy and security issues has been and will be undertaken, covering in particular German, Dutch, Danish and Swedish privacy laws, since these are the geographical locations of data controllers in the project, as use case leaders.

As regards anonymization techniques, data confidentiality, integrity and privacy will be assured when collecting and processing data. The information for each person contained in the release cannot be distinguished from a given number of other individuals whose information also appears in the release. Moreover, the pseudonymization of data is another method of ensuring confidentiality, according to the Article 29 Working Party Opinion on Anonymization Techniques and in relation to the EU General Data Protection Regulation. Where data are particularly sensitive (e.g. data using detailed personal narratives) then risks to confidentiality increase. In this case, participants will be carefully informed of the nature of the possible risks. This does not preclude the responsibility of the applicant to ensure that maximal anonymization procedures are implemented. A detailed description of the measures that will be implemented to prevent improper use, improper data disclosure scenarios and 'mission creep' (i.e., unforeseen usage of data by any third party), within the above-mentioned security protection strategy, will be provided as update of this deliverable.

The optimal solution will be decided by using a combination of different techniques, while taking into account the practical recommendations developed in the above-mentioned Article 29 Working Party Opinion on Anonymization Techniques. It should be noticed that, although pseudonymization approaches reduces the link ability of a dataset with the original identity of a data subject and is accordingly a useful security measure, according to both the above mentioned Article 29 Working Party Opinion on Anonymization Techniques and the General Data Protection Regulation, pseudonymization does not qualify as an anonymization technique, as it enables data to re-identify the data subject to which it refers. Therefore, pseudonymized data is personal data and, therefore, its processing must comply with the GDPR in full.

These techniques should adhere to certain requirements to comply with data protection and privacy-related legislation in the EU. The following set of requirements (among others) has been extracted from the GDPR and the Article 29 Working Party Opinion on Anonymization Techniques and will be the guidelines for security protection strategy drafting:

- User authentication: the system has to provide adequate mechanisms for user authentication.

- Limited access: the system must ensure that data is only provided to authenticated and authorized persons. The list of authorized persons to each dataset and their user privileges shall be restricted to the minimum necessary.
- Protection against unauthorized and authorized access: the records of an individual have to be protected against unauthorized access.
- Notice about use of data: the users should be informed about any access to their records.
- Access and copy users' own data: the system must provide mechanisms to access and copy the users' own data.
- Modification of the database: if an attacker breaks into the system, the system must detect modifications and inform the system administrator about this attack.
- Data protection by design and by default: privacy and data protection standards shall be taken into account from the outset of the project. Therefore, taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures which are designed to implement data-protection principles in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of the General Data Protection Regulation. Additionally, the controller shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.

7.2.4. Internal threats and human errors

Most organizations focus on data management risk from external threat, but most breeches occur from internal vulnerabilities. These can be thought of as part of the same risk continuum. This section looks at internal vulnerabilities and how to reduce them. There are two main types of internal threats:

- Security may fall victim to **human error**. For example, an employee may copy information from an entire database table into an email for troubleshooting purposes and accidentally include external email addresses in the recipient list.
- **Internal Attacks**. While internal accidents often compromise databases, willful attackers on the inside commit a large portion of database breeches. Many are disgruntled employees who use their privileged access to damage.

Most of these attacks came using the numerous outlets for data on the modern PC, including USB and Firewire ports, CD and DVD recorders and even built-in storage media slots. Combined with the fact that

storage space on portable devices has rapidly increased, business professionals can now use personal storage devices, such as USB memory sticks, IPods, digital cameras and smart phones, to remove or copy sensitive information either for malicious intent or personal gain.

Internal threat prevention

The implementation of a strong and flexible security policy is essential for MarginUp!. A security policy can provide rules and permissions that are understandable to both the employee of MarginUp! partner organizations and those implementing them so that personal data is prevented from leaving the office. MarginUp! policy is based on the security policies in the EU that are often enough if enforced to prevent such breeches, and are summarized in the following 5 points methodology:

Table 3. Internal threats prevention methodology

No.	Prevention Method	Implementation	Beneficiaries, Timeline
1	Data protection policies	Using national or local legal guidelines for data protection and privacy policies (DP)	Consortium, disseminated print and electronic document (Nextcloud), within first 6 months
2	Internal data protection policies	Written policies and procedures for all staff to sign in and agree to	Consortium, within first 6 months and with new coming employees
3	Clear staff role definition and responsibilities	Staff training, awareness and clear roles and staff responsibilities on data for access to data with checklists	Consortium, Regular monthly organizing trainings or integrating with internal meetings
4	Access control	Managing change in staff and have leave processes in place	Project coordinator, organizations leaders, when there is a change
5	Sanctions and audits	Disciplinary action for breach of DP and process guidelines by staff and threat of audits	Consortium, when a violation occurs

7.3. Data security as specified for Zenodo

Open results deposited in the Zenodo repository are safely stored for long time preservation. The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN, has defined a lifetime for the repository of the next 20 years minimum.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content.
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject based repositories.

8. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and deliverable of ethical guidelines (D8.2). Include references and related technical aspects if not covered by the former Informed consent to processing of personal data and is included during data collection. Sensitive data will be separated as soon as possible and kept secure.

In accordance with the General Data Protection Regulation 2016/679, the data controllers and processors are fully accountable for the data processing operations.

9. Bibliography

- Creative Commons License (2023). Discover the new CC Search. *https://creativecommons.org/licenses/* (Accessed 19 May 2023).
- European Commission. Guidelines on Fair Data Management in Horizon 2020. Version 3.0. 26 July 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- European Commission. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 Version 3.2. 21 March 2017, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- European Commission. IPR Helpdesk. Fact Sheet Open Access to scientific publications and research data in Horizon 2020.
- OpenAIRE Guides and Factsheets: https://www.openaire.eu/support.
  Open access (Digital Agenda site): http://ec.europa.eu/research/openscience/index.cfm.
- Information Platform for Max Planck Researchers – by MPDL. Research Data Management, https://rdm.mpdl.mpg.de/2022/04/04/data-management-plans-for-horizon-europe-and-erc-grants/.
- *https://enspire.science/data-management-plan-in-horizon-europe/* (Accessed 05 April 2023).
- European Commission. Grant Agreement number 10108208 – MarginUp!.
- MarginUp! Consortium Agreement.

ANNEX 1 - Data Inventory Table

Data set no.	Data set 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 UtilityWho outside of the project consortium might use the data?	Type of IP/ Protection sought	How will data be reuse	Ethical Issues? Y/N