Data management plan 1

WP7: COORDINATING, NETWORKING, AND DATA MANAGEMENT
TASK 7.2: DATA MANAGEMENT AND PUBLICATION STRATEGY

Deliverable D7.2

30 December 2024
Version 1

Dimitry Wintermantel, Amibeth Thompson, Felix Fornoff
University of Freiburg
Sara Leonhardt, Paula Prucker, *TUM*
Zehra Basaran, Georgios Kleftodimos, *CIHEAM-IAMM*
Georg Andersson, Maj Rundlöf, *ULUND*
Oliver Schweiger, *UFZ* / Nicola Gallai, *ENSFEA*
Martin Thorsøe, *AU* / Annie Ouin, *INPT*
Remco Ploeg, *WU* / Lynn Dicks, *UCAM*
Tom Breeze, *UREAD*

RestPoll
Restoring Pollinator habitats across European agricultural landscapes based on multi-actor participatory approaches

Prepared under contract from the European Commission
Grant agreement No. 101082102
EU Horizon Europe Framework Programme

Project acronym:	**RestPoll**
Project full title:	**Restoring Pollinator habitats across European agricultural landscapes based on multi-actor participatory approaches**
Start of the project:	Oct 2023
Duration:	48 months
Project coordinator:	Professor Alexandra-Maria Klein University of Freiburg, Freiburg, Germany www.restpoll.eu
Deliverable title:	Data management plan 1
Deliverable n°:	D7.2
Nature of the deliverable:	Report
Dissemination level:	Public
WP responsible:	WP7
Lead beneficiary:	ALU
Citation:	Wintermantel, D., Thompson, A., Fornoff, F. *et al.* (2024). *Data management 1*. Deliverable D7.2. EU Horizon RestPoll Project, Grant agreement No. 101082102.
Due date of deliverable:	Month 15
Actual submission date:	Month 15

Deliverable status:

Version	Status	Date	Author(s)
1.0	Submission to EU	30 Dec 2024	Wintermantel [ALU-FR], Thompson [ALU-FR], Fornoff [ALU-FR], Leonhardt [TUM], Prucker [TUM], Basaran [CIHEAM-IAMM], Kleftodimos [CIHEAM-IAMM], Andersson [ULUND], Rundlöf [ULUND], Schweiger [UFZ], Gallai [ENSFEA], Thorsøe [AU], Ouin [INPT], Ploeg [WU], Dicks [UCAM], Breeze [UREAD]

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

Table of contents

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

Preface

This plan outlines the workflow for data management, including the creation, curation and sharing of data. The responsibilities of the file managers are outlined, along with the guidelines for creating and curating data on the RestPoll drive.

1. Summary

This plan outlines a comprehensive workflow for managing research data within the RestPoll project, from collection to publication. The RestPoll drive (a Microsoft SharePoint page) is used to store and exchange data within the project. On this drive data are accessible to all RestPoll members. Sensitive personal data are not stored on the RestPoll drive but on secure institutional drives.

Central to the plan is the RestPoll data dictionary, where all data files with their respective file managers are listed and all variables across all RestPoll data files as well as key concepts are defined (see 'Data dictionary' below).

Data file managers are tasked with preparing structured files from templates for data entry, ensuring these files include relevant metadata. Common protocols for data collection will reduce risk of data not being consistent and safeguard data quality. Data collection procedures will also be calibrated before start and data files may include some features that facilitate correct data entry (see 'Structure of spreadsheet files' below). After data entry, a second person should check a random selection of entries to validate quality.

RestPoll will generate data on pollinators, pollination success and land use metrics in and around the case-study areas. This will be mostly in form of numeric data, but also image, geospatial and text documents. In addition, RestPoll will possibly make use of existing data on the above or possible covariants. All data will be stored along with rich metadata. For reused data, the metadata documents their provenance. Data will typically be stored in excel files or GeoPackages (in the case of spatial data). Before publication, Excel files will be transformed to ODS or CSV files. Photos, in a lossless compressed format (which allow for reconstruction of compressed data without loss of information) such as compressed TIFF or PNG, will be created to document field work and possibly for identification of pollinators. Other formats include RAW, BMP, or GIF. Recordings of workshops or interviews may be retained, with consent from participants, as .WAV or .mp4 files or similar, although they may only be stored temporarily for transcription purposes and then deleted. Transcripts, protocols, surveys / questionnaires and other qualitative data will be saved as .txt or .doc files. Other data types resulting from this project include, for example, R-files (.R, .Rmd) and R-packages from the open-source software R.

The expected size of the data generated or re-used by RestPoll will depend to a considerable extent on how many photos and recordings will be created, which in turn will depend on whether photos of pollinators and plants will be stored for identification. The RestPoll drive may require storage for 200 GB to 1 TB.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

2. General workflow

Below is the proposed workflow for data preparation, creation, and curation, but is subject to change based on feedback during the progress of the project. Members who would like to suggest changes or feedback, please contact Dimitry Wintermantel (dimitry.wintermantel@nature.uni-freiburg.de) and Amibeth Thompson (amibeth.thompson@nature.uni-freiburg.de).

1. Check the RestPoll **data dictionary** for current terms and variables.
2. **Preparation of protocols** that specify what variables should be recorded, using variables already existing in the data dictionary or using new unique variables; variables typically include date of data collection (or action), site identity, observer(s) etc.
3. **Update** of the RestPoll **data dictionary** by adding the data file name (in the 'Data_files' sheet) and new (i.e. previously undefined) variables (in the 'Variables' sheet; Fig. 1).
4. **Preparation of files** by the respective file managers (each file has one file manager). The structure of the file and potentially also some mock data (that is clearly indicated to be mock data and that is deleted after some real data has been entered) should facilitate correct data entry. The files should contain metadata both on the entire file (see 'General metada in README sheets/files') as well as on individual variables (see 'Structure of spreadsheet/qualitative files') but no sensitive information such as personal data of land owners.
5. Data **collection.**
6. Data **entry. Validation** of random samples of data entries should be carried out by a second person to verify that data has been entered correctly.
7. Data cleaning/correction and documentation of the process.
8. Data **analysis** and documentation of the process (e.g. by using detailed methods descriptions or syntax-based files, where meaningful function and object names as well as comments facilitate readability).
9. Typically, at the time of **publication** (see publication policy in Deliverable 7.1) of peer-reviewed research, a copy of the data used for the publication shall be saved in an **open-source format** (excel to ODS or CSV and word to ODT or TXT). Simple text-based file formats such as CSV and TXT are likely most enduring, but ODS and ODT files allow keeping multi-sheet files easily together. Therefore, data file managers should decide between the two forms (CSV/TXT or ODS/ODT). These files shall be stored in subfolders (see 'Naming conventions' below for how subfolders and files should be named).
10. These shall be saved in subfolders named after the original file, the date of editing/transforming, and the person who transformed the data. These should then be **uploaded** to the **Zenodo** repository and a platform more targeted to the data (see 'Repositories for publication').

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

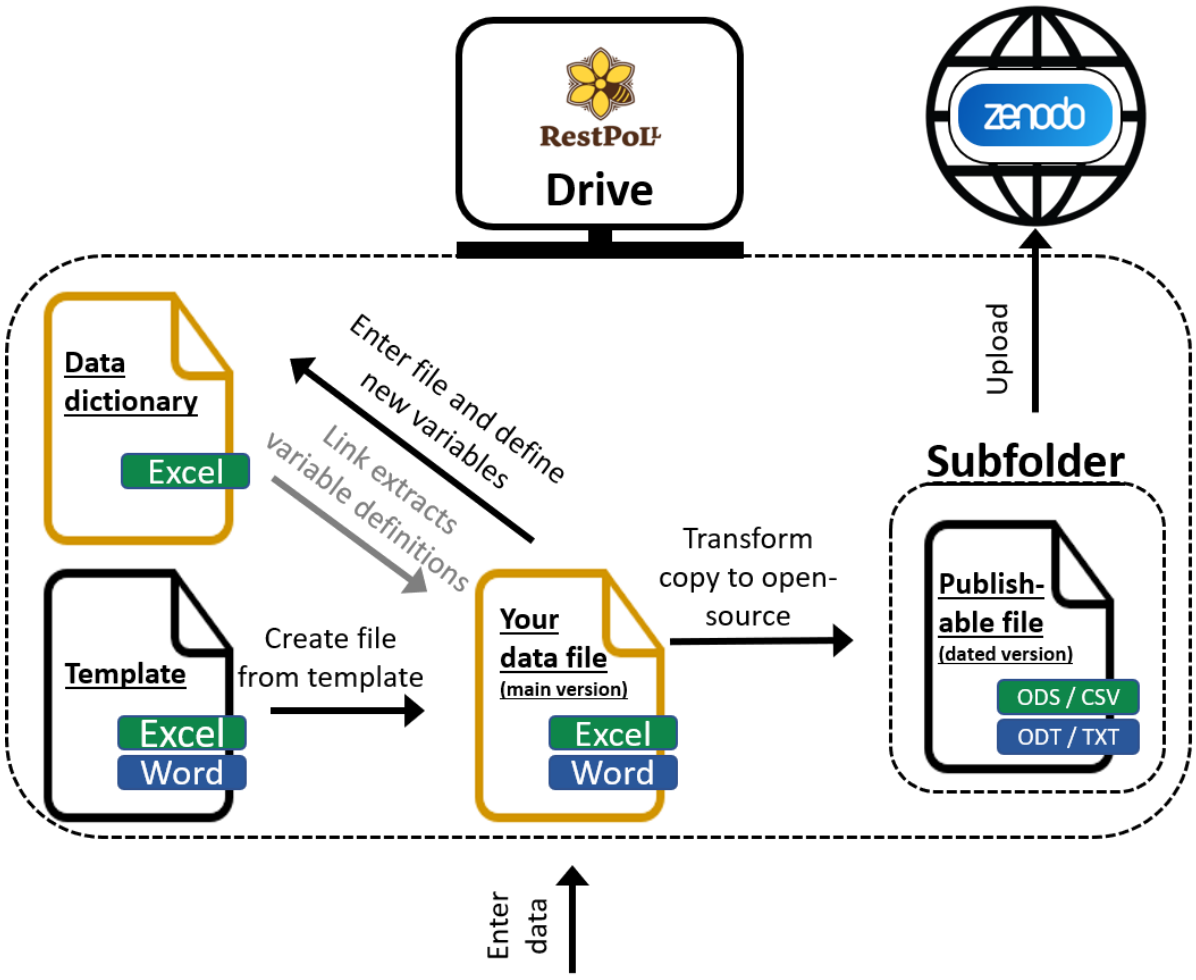


Figure 1. Overview of the data management workflow. On the RestPoll drive, which is accessible to all RestPoll members, a file manager creates a data file (labelled 'Your data file') from a template, which is linked to the data dictionary, where all variables are to be defined. Data will continuously be entered in the data file. Eventually, a copy of the data file will be transformed to an open-source format, saved with the date added to the file name in a subfolder of the folder where the main data file is stored on the RestPoll drive and published on Zenodo and one or more platforms targeted to the type of data.

3. RestPoll drive and folder structure

Microsoft SharePoint will be used to store and manage data. The "RestPoll Data" SharePoint site consists of four document libraries: *Deliverable & Milestones*, *Library*, *Public Data*, and *Data* (Figure 2). The *Deliverable & Milestone* library will be linked to the final version of the respective document/file/folder from the WP folder (in the *Data* library). This folder will be assessable for all RestPoll members. The *Library* library will be for storing and sharing identification keys for insects and plants and relevant literature for within and across Tasks. This library will be assessable for all RestPoll members. The *Public Data* library will include folders that contain the final files for any public (published) data (i.e. manuscripts, reports, etc.). This will be linked from the respective "Public Data"

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

folder in the *Data* library. The *Data* library will store all of the raw and final data collected within RestPoll (details below). This library will only be accessible by Data File managers and other RestPoll members responsible for data creation or entry.

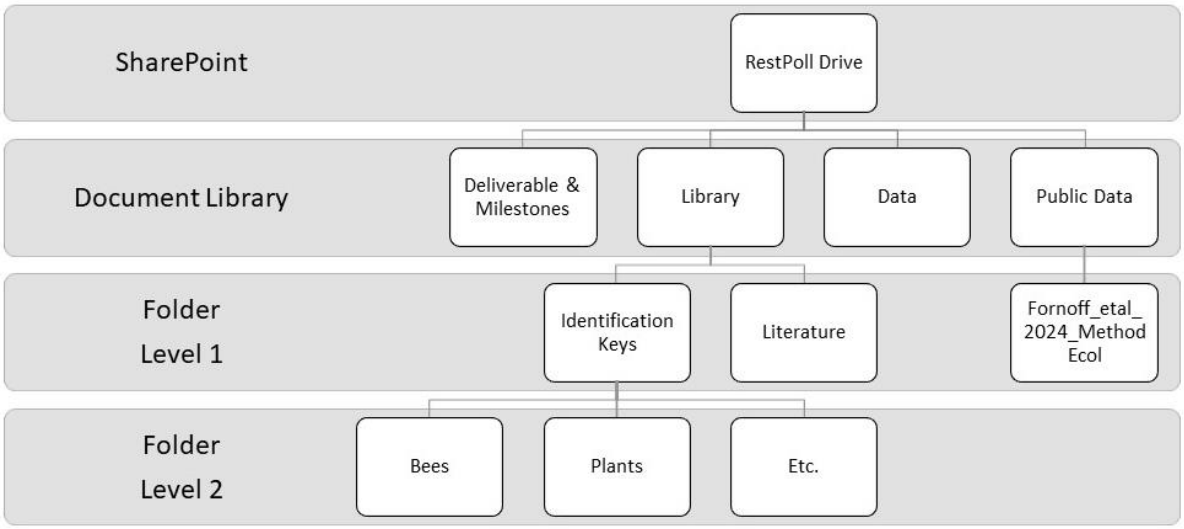


Figure 2: Overview of the document libraries on the "RestPoll Data" Sharepoint site.

The *Data* library will be the main data storage and management location for data collected within the realm of the RestPoll project and the Data Dictionary. Each Work Package (WP) will have its own folder for storing data. The first three levels of folders will be identical for all WPs (Figure 3). Within each WP folder (level 1) will be a folder for each *Task* (level 2). Within each *Task* folder, there will be a folder for storing *Raw Data*, for aggregated *Data*, the *Deliverable & Milestones*, and for the *Public Data* (level 3). Additional folder levels may be created by the Data file managers of the affiliated WP / Task to help organize the stored data. When possible, please abide by the Naming Conventions (see section 4), however there is flexibility for the *Raw Data* folder. See Figures 4 and 5 for examples for WP 1- Tasks 1.1 and 1.2.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

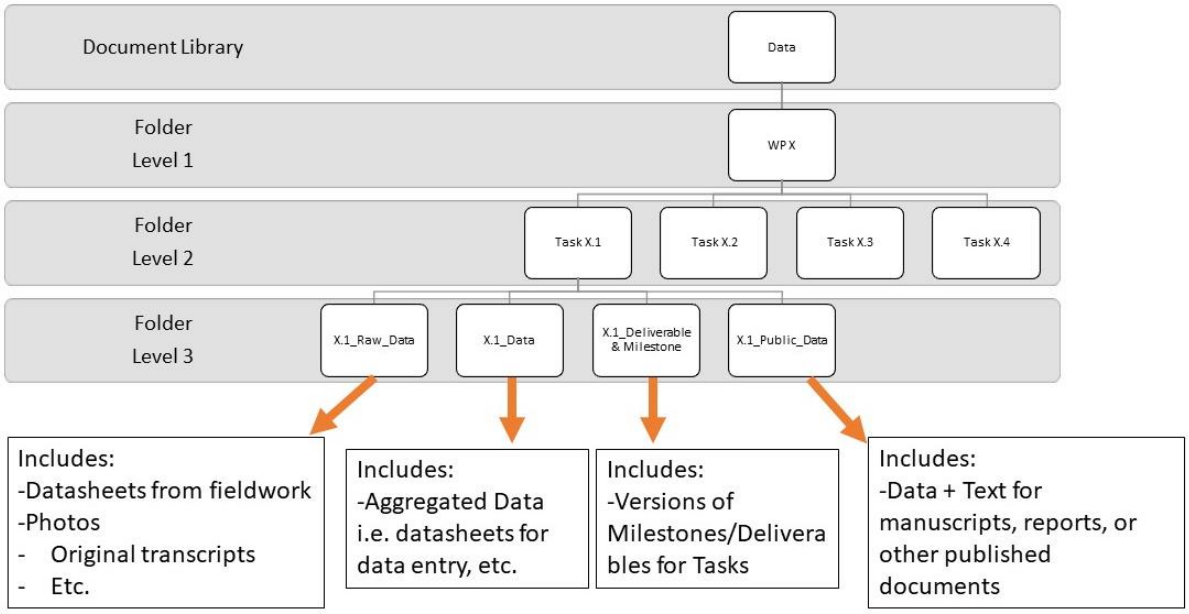


Figure 3: Overview of the folder structure for each Work Package in the *Data* library.

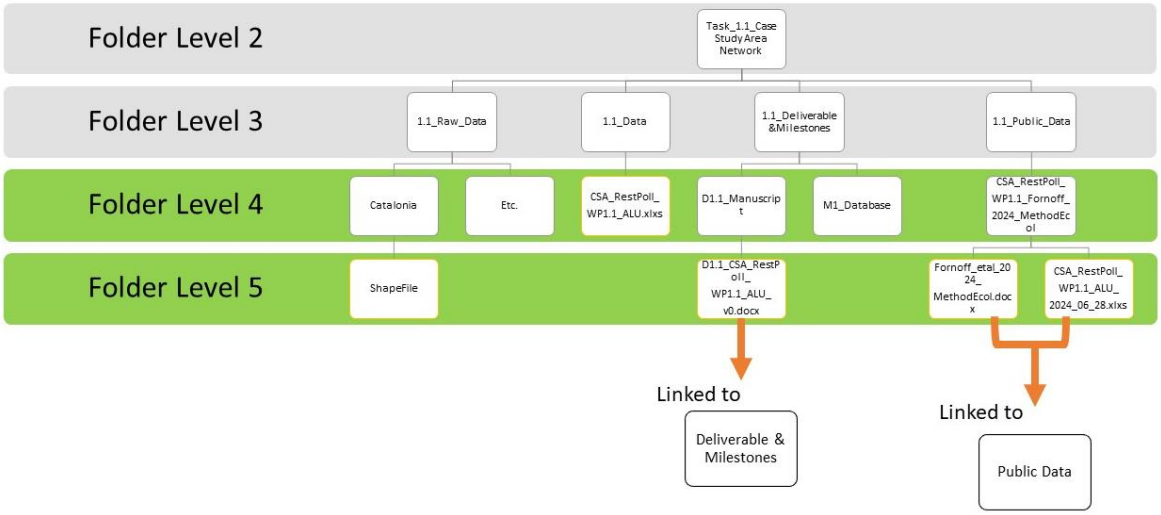


Figure 4: Example of additional folder level for Task 1.1. An additional folder has been added for each case study areas in *Raw Data*, whereas the *Data* folder contains one excel-sheet with all of the aggregated data.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

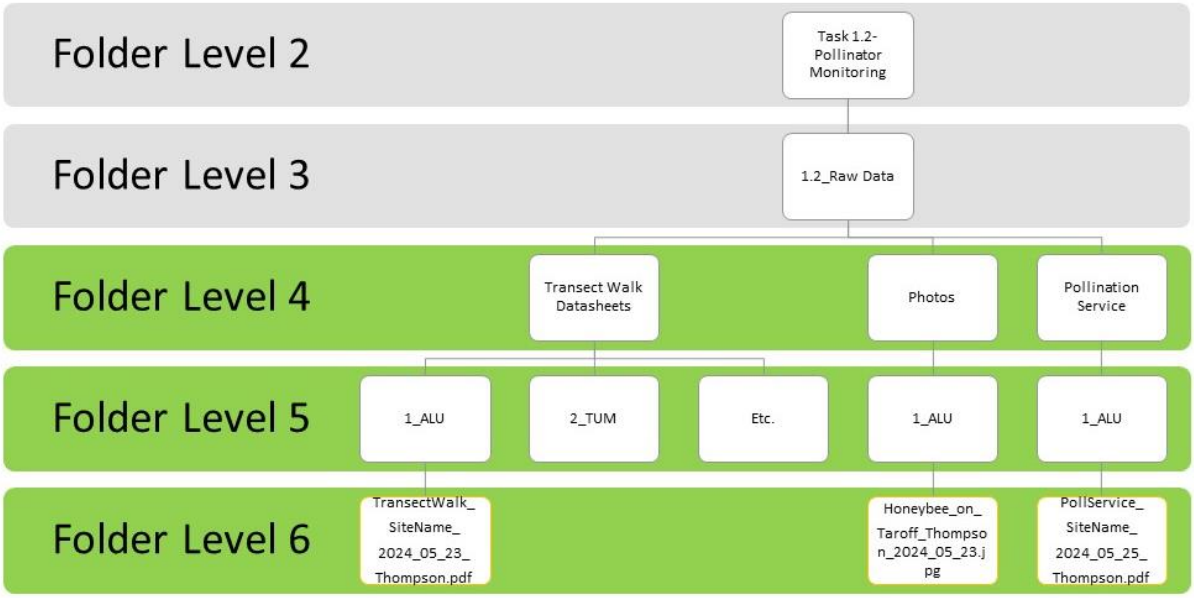


Figure 5: Example of additional folder levels for Task 1.2. Folder level 4 contains the different types of data that will be collected (*Transect walk*, *Photos*, and *Pollination services*) and then level 5 are the folders for each *case study area*. Within the *case study area* folders, it is further divided for the different *sites and dates* when the data was collected.

4. Naming conventions

Files, sheets and variable names should use underscore (_) rather than spaces or hyphens to separate words and start with letters rather than numbers to avoid potential issues in some programming languages.

4.1. VARIABLE NAMES

Variable names should be simple and descriptive. They start with an uppercase letter and are followed by lowercase letters. Acronyms, used only if widely understood and accepted (e.g. DNA), are typically spelled in uppercase letters or how they are spelled in standard English (e.g. DNA or cDNA). Different words are separated by an underscore (_) e.g. “Bombus_abundance”.

4.2. FILE NAMES

Relatively simple, descriptive file names will allow other RestPoll members to easily find the data files they need. Files should be named using this convention:

[Data title]_RestPoll_WP[WP number. Task]_[Partner acronym] e.g. “Pollinator_survey_RestPoll_WP1.2_ALU.xlsx”.

In rare instances, it may make sense to store data in a format other than the tidy data format (i.e. for overview rather than analysis). In this case, the file name should contain “_overview” after the data title e.g. “Sites_overview_RestPoll_WP1.1_ALU.xlsx”.

Funded by the European Union

This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.

For Deliverables and Milestones, please include the number before the "Data title" and the version number (with lowercase "v") at the end. For example:

D/M[Deliverable/Milestone number]_[Data title]_RestPoll_WP[WP number. Task]_[Partner acronym]_[version number] e.g. "D7.1_Communication_and_management_plan_RestPoll_WP7.2_ALU_v0.docx"

Updated version numbers should be assigned after each revision. In principle, the versions should be: v0= first draft [if sent out for feedback before first submission], v1= first revision and/or submitted document to the EU portal, v2= second revision with feedback from EU or updates, v3= third revision/update, etc. Versions of the documents between revisions can be identified as X.1, X.2, etc., if there is a need for additional documentation of changes. If necessary, reasons for or highlights of major changes should be explained in a new section at the end of the document.

Data files on the RestPoll drive generally do not require version date (as there should be only one main version that is constantly being updated). Versions of specific dates may be saved in subfolders, especially datasets that are transformed to an open-source format and published on Zenodo (and other repositories) should contain the version date (i.e. the date the file has been edited) at the end of the name in the format **[Year]_[Month]_[Day]**.

In this case the naming convention is

[Data title]_RestPoll_WP[WP number. Task]_[Partner acronym]_[Version date] so, e.g. "Pollinator_survey_RestPoll_WP1.2_ALU_2026_05_28.xlsx".

When multi-sheet files (e.g. in excel) are being transformed to a single-sheet format (e.g. CSV or TXT), the individual files should end in the sheet name. In that case the naming convention is therefore:

[Data title]_RestPoll_WP[WP number. Task]_[Partner acronym]_[Version date]_[Sheet name], so, e.g. "Pollinator_survey_RestPoll_WP1.2_ALU_2026_05_28_README.csv" or "Pollinator_survey_RestPoll_WP1.2_ALU_2026_05_28_Pollinators.csv"

File names for photos are as followed:

[Descriptive title]_[Photographers surname]_RestPoll, e.g. "Honeybee_on_willow_flower_Fornoff_RestPoll.jpg". Optionally, they may also inlcude a date in the format year_month_day, e.g. "Honeybee_on_willow_flower_Fornoff_RestPoll_2025_04_16.jpg". More information on the photos should be included in the metadata of the file (i.e. GPS location, time stamp, etc.).

File names for public (publishable) data should include the first authors name, the year of submission, and the journal of submission. For example,

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

[First author's last name]_etal_[Year]_[Journal] e.g. "Fornoff_etal_2024_MethodEcol.docx"

4.3. PUBLIC DATA FOLDER NAMES

Dated files (i.e. files with a version date rather than the file that is constantly updated) to be used for public data shall be stored in subfolders within the respective Task folders. These subfolders shall be named following this convention: **[Data title]_RestPoll_WP[WP number. Task]_[First author of publication]_[Year of publication]_[Acronym of Journal]** so e.g. "Pollinator_survey_RestPoll_WP1.2_Doe_2027_STOTEN". This may of course be amended if the data were not published in a journal.

The main folder structure is created by Amibeth Thompson. Changes to this are to be discussed with her.

5. General metadata in README sheets/files

For each data file, general metadata shall be provided in form of a README sheet/section/file that is either part of the data file or a separate file linked to the data file. General metadata include a description of the data, the name of the file manager, his/her email address, the project name (RestPoll), work packages and partners involved in the creation of the file, countries where data were collected, date of creation, links to protocols used for data generation and optionally keywords to be entered in Zenodo (or other repositories) when publishing the data and links to publications using the data (the latter applies mostly to open-source files for publication and long-term storage).

6. Data dictionary

The data dictionary is central to the data management in RestPoll as this is where all variables used in RestPoll as well as important concepts are defined. It will be located in the *Data* library, outside of any WP folder. It also provides an overview of the data files and their data file managers. The data dictionary contains four sheets

1. a **README sheet** containing **metadata** on the file
2. a 'Data_files' sheet where data files along with their file managers and further information are listed.
3. a sheet named 'Variables' that defines **all variables** that are used in **any RestPoll dataset** and
4. a sheet named 'Key_concepts', which contains definitions of important **concepts** that are not variables.

The data dictionary shall enable quick comprehension of data sets and easy integration of different data sets. Any variable found in a RestPoll dataset should be found in the 'Variables' sheet of the data dictionary exactly once together with its definition, i.e. all variables should be included, and no variable should be included twice under different names. Therefore, when creating a data file, one shall ensure that variables are named as in the data dictionary. If the dataset contains a variable not yet found in the data

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

dictionary, the data dictionary shall be amended by adding the variable including its definition, the unit, the person who added the variable ('Added_by') and the date of adding the variable ('Date_added').

Definitions can be updated only with consensus from the original definers and the file managers of all affected data files, listed in a separate column ('Files_used_in'). Old definitions are archived under 'Former_definitions', separated by semi-colons if there are multiple. Each entry includes who added the definition originally and who modified it along with the dates. For instance, if 'Variable' was modified twice, its entry might read: "An element, feature, or factor that can vary or change (defined by Dimitry Wintermantel on 2024-01-13, modified by Felix Fornoff on 2024-01-30); A column heading (defined by Felix Fornoff on 2024-01-30, modified by Jane Doe on 2024-02-13)". The new definition should be under 'Definition', Jane Doe should be under 'Added_by' and the date she added the new definition (2024-02-13) under 'Date_added'. If the name of a variable is changed, former variable names shall be provided in an analogous way under 'Former_variable_names'.

Along with this information, the format of the variable, the unit (enter 'none' in the case of unit-less variables) and an example shall be provided. The format refers to the data type, e.g. numerical, or character. Factors are like characters except that they have defined values (hereafter value options), which are noted to the right-side of a black bar. 'Factor_fixed' refers to a factor where no additional value options can be added, whereas for a 'Factor_editable' additional value options can be created.

The data dictionary has a 'Filter' column that allows for filtering for subsets of variables to facilitate quickly checking if a certain variable exists already. The 'Filter' variable is an example of a 'Factor_editable' as more value options can be entered over time. In the data dictionary, in the column 'Data_files_used_in', it shall be specified in which data files a particular variable is used in, which can also facilitate quickly finding relevant variables. If anyone wants to enter additional columns in the Data dictionary, please see the note on entering columns under 7.2 and inform Dimitry Wintermantel (dimitry.wintermantel@nature.uni-freiburg.de)

7. Data file structure

7.1. QUANTITATIVE DATA FILES

Most data shall be stored in excel files that contain:

1. A **README sheet**
2. An '**ACTIVITY**' sheet that tracks changes to the file. People modifying the file enter here shortly what they did, e.g. 'Data entry'. File managers can delete this sheet when creating the file if they decide this is not necessary. A 'Variables' sheet that **extracts definitions** and related information of the **variables** used in the data sheets from the **data dictionary**
3. (A) **data** sheet(s)

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

Data file managers shall adhere to the following procedure for creating an excel file:

1. **Create a file from** the template ('Data_file_template_RestPoll.xlsx', see Figure 6), **rename it** (see 'Naming conventions') within the correct folder.
2. **Enter** the **data file** in the 'Data_files' sheet of the data dictionary.
3. Enter general **metadata** in the **README** sheet as indicated by the template and described under 'General metadata in README sheets/files'.
4. Create data sheets to the right of the Variable sheet. These shall generally be organized according to the **tidy data principles** (https://www.jstatsoft.org/article/view/v059i10). Tidy data sets are arranged so that each variable forms a column, each observation forms a row, each value must have its own cell, and each type of observational unit forms a table (https://r4ds.had.co.nz/tidy-data.html, Wickham 2014). Data collected at **different levels** (also named observational units) should be stored in **different sheets**. Different levels may for example be case-study area (CSA), farm and site. These sheets shall be named after these levels, e.g. 'CSAs', 'Farms', 'Sites'. Data should only be included in the sheets where they are collected so that only identifying/defining variables such as Farm (identity) overlap between different data sheets (in this case 'Farms' and 'Sites'). The column combination that uniquely defines an observation should be placed in the beginning of the sheets and be highlighted in light grey. For example, if 'Site' and 'Date' make an unique observation, these two columns should be marked grey. No cells are combined (merged) and all variable names are written in a single (top) row.
5. Ensure that the **variables** in the data file are **defined** in the **data dictionary**, if not define them there (in the data dictionary, not the data file). The "Filter" column of the data dictionary should allow you to quickly see if a related variable exists already. If two variables with the same name are entered they are highlighted red (as this should not happen).
6. In the data file, **copy the variable names** and **paste** them in a **transposed** way in the first column ('Variable') of the 'Variable' sheet. Repeat subsequently down the rows of the first column for all sheets. The definitions should appear. Variable names that appear more than once are highlighted yellow. These are typically variables that can be used for joining data sets from different sheets.
   Optional: **Correct data entry** can be promoted by using the **data validation feature** in excel (or validity feature in Open Office). The feature (https://support.microsoft.com/en-us/office/apply-data-validation-to-cells-29fecbcc-d1b9-42c1-9d76-eff3ce5f7249) allows restricting the format (such as date, decimals or whole numbers) that may be entered in certain cells as well as the range of permissible values (e.g. only values greater than zero). To restrict factor values, select the column (excluding the header) in the 'Variable' sheet. Then, navigate to Data > Data Validation, choose 'Allow: List', and select values listed to the right of the black bar. This will create a drop-down menu from which people entering data can choose from. Make sure that for 'Factor_editable' you

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

include a number of empty cells on the right side of the last value added as more values may be added (in the data dictionary that are then transferred to the data file). A less restrictive but informative aid for data entry is an input message that details what values can be entered.

7.2. IMPORTANT NOTE ON INSERTING COLUMNS:

When inserting a column in Excel, it automatically inherits the formatting, including data validation features, of the column immediately to its left. This behavior can result in unintended consequences, such as incorrect data validation settings or formatting. To avoid this, insert columns in one of these two ways:

1. **Copy and insert a standard column**. Navigate to a column without special formatting or data validation features; higlight this and insert it where needed.
2. **Insert new column and then manually clear formatting and data validation**:
    a. **Clear Formatting**:
        i. After inserting the new column, select it by clicking its header.
        ii. Navigate to the **Home** tab.
        iii. In the **Editing** group, click on **Clear**.
        iv. Choose **Clear Formats** to remove all formatting from the selected column.
    b. **Remove Data Validation**:
        i. With the column still selected, go to the **Data** tab.
        ii. Click on **Data Validation** in the **Data Tools** group.
        iii. In the dialog box that appears, click **Clear All**.
        iv. Click **OK** to remove any data validation rules from the column.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

3.

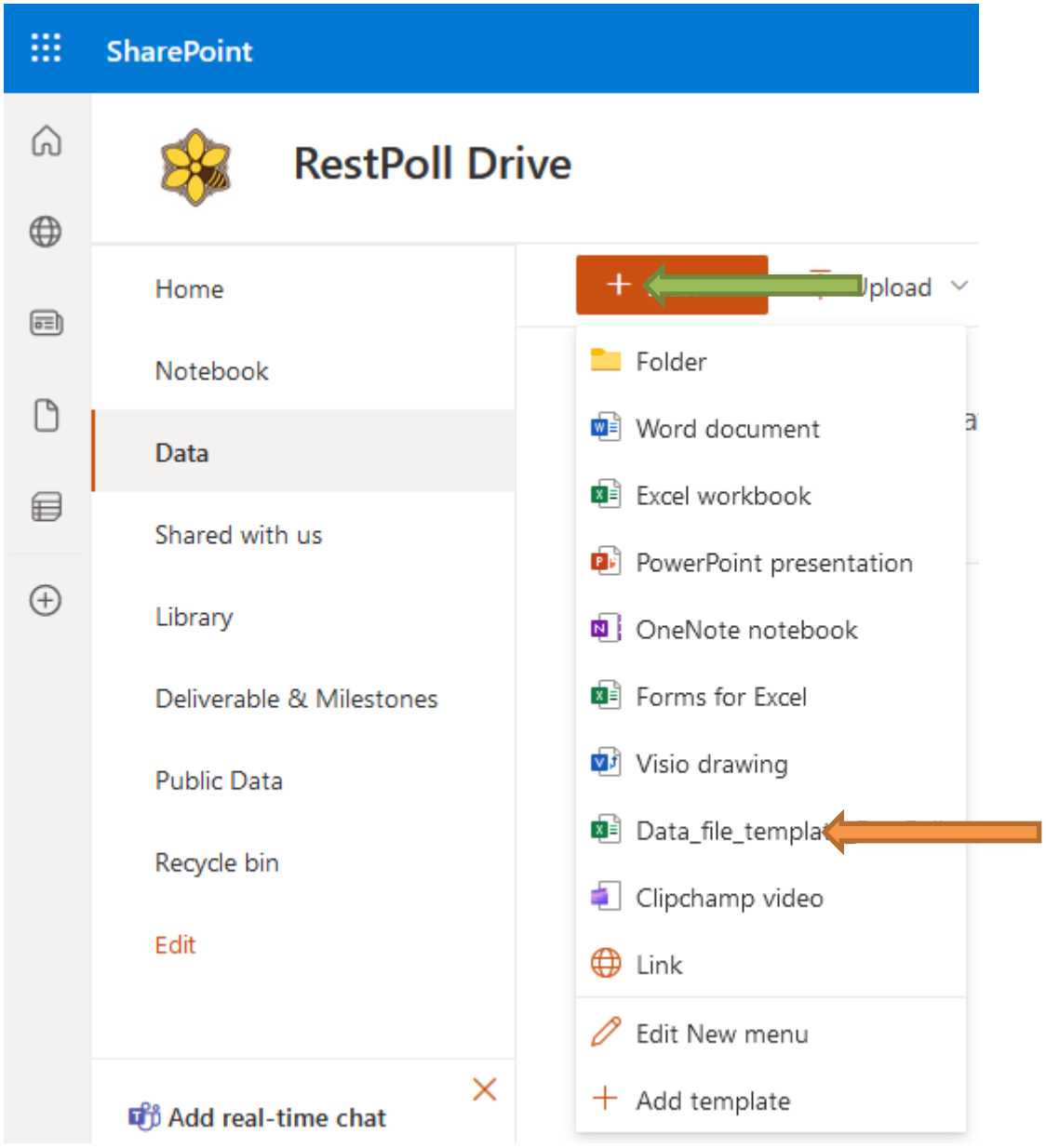


Figure 6: Screenshot of RestPoll drive when creating a new excel file from the RestPoll template. In the folder you want the new document, click " + New" (green arrow) and then "Data_file_template_RestPoll" (orange arrow).

7.3. QUALITATIVE DATA FILES

Some data will be collected for qualitative analyses purpose, as for example, the transcription of interviews. These data will be integrated to a word document. The document will include a summary at the beginning, which includes a list of WP(s) involved, a description of the interview, the dates and the objective. The document will be divided by title and subtitle. The title will describe the objective, the period and the interviewer. The subtitle will recall the codename of the interviewee and the date.

8. Data storage and backup

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

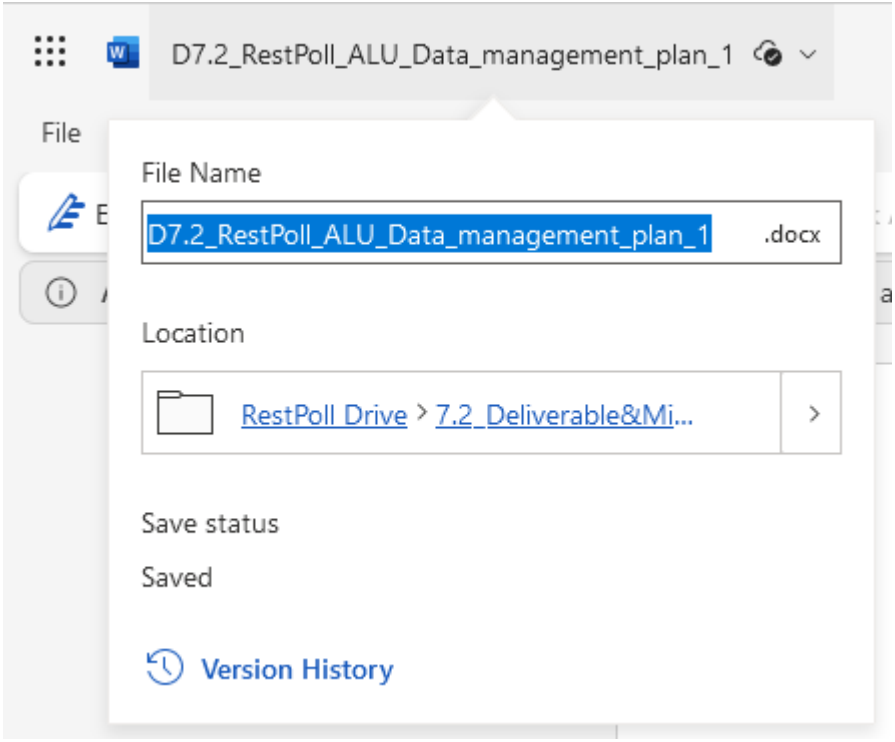
Backups of data will be done regularly by the file managers, on hard drives and cloud services. Backup storage shall be password-protected. Backups of data on the RestPoll drive will be done on the bi-weekly basis. Long term storage is procured by the local universities and on **Zenodo**.

8.1. TRACK-CHANGES AND VERSION CONTROL

When working on a text document, it is suggested that track-changes be used when working with multiple partners or revisions. Otherwise, "Version History" can be used to restore deleted or changed text. When working on datasheets, mistakes can occur when entering data or rearranging data (i.e. mixing of data across rows or columns).

In the case the files should be restored using "Version History", the follow steps should be taken by the file manager:

1. Download the current version of the document to save potential new additions not in the older version.
2. Open the document. Next to the file name click the "V" button, where a drop-down menu will appear. Click "Version History."



3. Here you will have multiple versions based on different users and dates when data has been entered. Restore or create a copy of the version of the document before the mistake has been made.
4. Add the data or information that was added after the mistake was made.

9. Repositories for publication

Data will be published in Zenodo and in repositories targeted to specific data such as GloBI, GBIF and the EU Pollinator Hub.the EU Pollinator Hub.

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

9.1. ZENODO

Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. All data should eventually be published here. Data should be published in the "RestPoll Horizon" Community (https://zenodo.org/communities/restpoll).

9.2. GLOBI

Global Biotic Interactions (GloBI) is an open-access repository focused on interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host). Please upload data here that involves two potential interactors, such as plants and pollinators.

9.3. GBIF

The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure funded by the world's governments. It provides open access to data about all types of life on Earth.

9.4. THE EU POLLINATOR HUB

The EU Pollinator Hub is an open-access interactive platform that centralizes and integrates data related to pollinators, bees, beekeeping, and other relevant data for the agricultural sector.

10.Responsible contact people

General data management: Dimitry Wintermantel (dimitry.wintermantel@nature.uni-freiburg.de)

Folder structure on RestPoll drive: Amibeth Thompson (amibeth.thompson@nature.uni-freiburg.de)

In addition, each data file has a ***data file manager*** with the following duties:

- Create file potentially with features that facilitate correct data entry
- Update data dictionary to include the file and all of its variables (make sure that only new variables are being added)
- Validate that data are being entered correctly, correct and inform people entering data if needed
- Ensure that file is regularly backed up and restore correct version in case of errors (see 'Data storage and backup')
- Inform all data-owning partners about requests to use/see the data
- Remind data users to put data in open-source format on Zenodo and a subfolder on the RestPoll drive when it is time to publish the data

10.1. DATA OWNERSHIP

In accordance with the Grant Agreement (Article 16), results are owned by the beneficiaries that generate them. Two or more beneficiaries will own results jointly if they

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

have jointly generated them and it is not possible to establish the respective contribution of each beneficiary, or separate them for the purpose of applying for, obtaining or maintaining their protection. In case of joint ownership, the joint owners have the obligation to conclude a joint ownership agreement.

The beneficiaries may transfer ownership of their results, provided this does not affect compliance with their obligations under the Agreement and must inform the other beneficiaries with access rights of the transfer at least 45 days in advance (or less if agreed in writing), unless agreed otherwise in writing for specifically identified third parties including affiliated entities or unless impossible under the applicable law.

In each file, the partners that own the data are listed in the README sheet. The data owners shall

- decide whether the data can be used for certain studies or shared with people outside of RestPoll
- suggest co-authors for publications using the data

This information will be used to create the results ownership list (ROL), which is a list of owners of the results generated by the action must be provided to the granting authority at the end of the action in the final reporting. When ownership of the results is not clear, all potential owners will be listed.

11. FAIR data

To ensure our data is FAIR, we share our data with existing infrastructures like GloBI, GBIF, and the EU Pollinator Hub, along with the “RestPoll Horizon” community within Zenodo.

11.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

To enhance data findability, each dataset in the RestPoll project will be assigned a unique, persistent identifier (DOI) upon deposition in the Zenodo public repository. This ensures that data sets can be reliably located and cited. Metadata accompanying these data sets will be rich and descriptive, including but not limited to, the dataset title, creator(s), publication date, summary, keywords, and geographic location (if applicable). We will adhere to disciplinary and general metadata standards to ensure consistency and quality. This metadata will be structured to make them easily findable by search engines. Additionally, the naming convention outlined above will facilitate intuitive searching and categorization of data sets.

11.2. MAKING DATA ACCESSIBLE

Once found, the person needs to know how to access the data. Both data and metadata should be retrievable by the identifier. The protocol for accessing the data is free, open and universally implementable.

11.3. MAKING DATA INTEROPERABLE

Funded by the European Union

This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.

The data, metadata, and documentation adhere to disciplinary standards. Additionally, interoperability is ensured through the above described comprehensive data dictionary, which will provide clear definitions and units for each variable, promoting uniform understanding and use across different data sets.

Files shall be made publicly available in open-source formats (typically ODS/ODT, CSV/TXT or geopackages). README sheets/files with metadata are provided.

11.4. INCREASE DATA RE-USE

The data shall be released under the Creative Commons license (CC0) as data cannot be copyrighted (Benichou et al. 2023). It will be linked via a DOI with publications detailing the provenance of the data sets. Manuscripts will be released under the latest Creative Commons BY license (CC-BY). All data will be made publicly available at the time of publishing peer-reviewed articles or 2 years after completion of the RestPoll project.

12. Other research outputs

Protocols will be publicly available on the RestPoll website. Teaching, training, and outreach materials will be publicly available on the RestPoll website.

Codes for analysis (R scripts, R packages, etc.) will be uploaded, linked to, and maintained on the RestPoll GitHub.

Insect specimens will be collected according to the Pollinator and Botanical Protocol (D1.2) and preserved by pinning according to the guidelines by Schauff 1986 (pg. 32). Specimens will be labeled with regional data (location, date, collector, flower visited), along with the identification (Order, Family, Species, Genus). The long-term preservation of the collections will be according to the affiliated institute's regulations or requirements, most often being preserved in a natural history museum. If an institute does not have a protocol or requirement already in place, we will provide an exemplary protocol that they can implement and adjust accordingly.

Photos will be uploaded to iNaturalist. This platform provides the posted content under the Creative Commons Attribution Noncommercial license (CC BY-NC). iNaturalist will be used to help with plant or inset identification that cannot be resolved by RestPoll members and will also be used to store pictures and make them publicly available. A RestPoll project has been created to link observations together (https://www.inaturalist.org/projects/restpoll). Instructions for uploading observations can be found here.

13. Allocation of resources

Individuals responsible for data management have been listed above. Costs for the RestPoll drive will be covered out of ALU's designated budget within the RestPoll grant.

14. Data security

To safeguard privacy and comply with the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679), sensitive data, such as contact details of land owners, will not

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

be stored on the RestPoll drive. Such information will be stored in secure, restricted-access locations by the responsible partner organizations.

For long-term accessibility, non-sensitive data will eventually be stored in the Zenodo repository and on the RestPoll website.

15. Ethics

Most of the ethical considerations relate to the activities in WP3, WP4 and WP5, in which interaction with stakeholders and collection of data representing perceptions, preferences and other and other forms of personal information. However, also WP1 and WP2 will rely on e.g. contact details to farmers and other land owners. The project will collect minimal personal data from participants, such as name, gender, email, and location. This data will be used only with consent for project activities and stored locally in a restricted access, access-controlled by the LL leader in each country. RestPoll will adhere to the "do no harm principle" and "anonymization principle" when analyzing data and producing outputs such as publications, networking reports, and communication supports. Lead institutions will be responsible for gaining ethical approval for gathering specific data sets of these types, when it is required. RestPoll confirms applying the ethical guidelines of the Horizon Europe project and all relevant EU/national legislation, international conventions, and declarations. This includes the Charter of Fundamental Rights of the EU, the European Convention for the Protection of Human Rights and Fundamental Freedoms, and GDPR.

In accordance with the GDPR, the cornerstone of our ethical practice is the informed consent from our informants. We will ensure that all participants clearly understand the purpose of the research and participate voluntarily, that there are clearly identified data controllers (i.e. person who is handling the raw data) and that, where possible, there are defined mechanisms for withdrawing data prior to analysis. Any publications or dissemination of project data will respect the participants' right to privacy, thus any request for deletion of data will be promptly honored.

No personal data will be included in published materials without explicit, informed consent from the individuals involved. To prevent the disclosure of personal information, all data containing personal information will be anonymized before storing and sharing among partners or publication online in accordance with the GDPR. Further, sensitive data will be securely saved from other data and not be shared between partners but stored separately by each partner organization.

16. Other issues

The data management procedures are as outlined above.

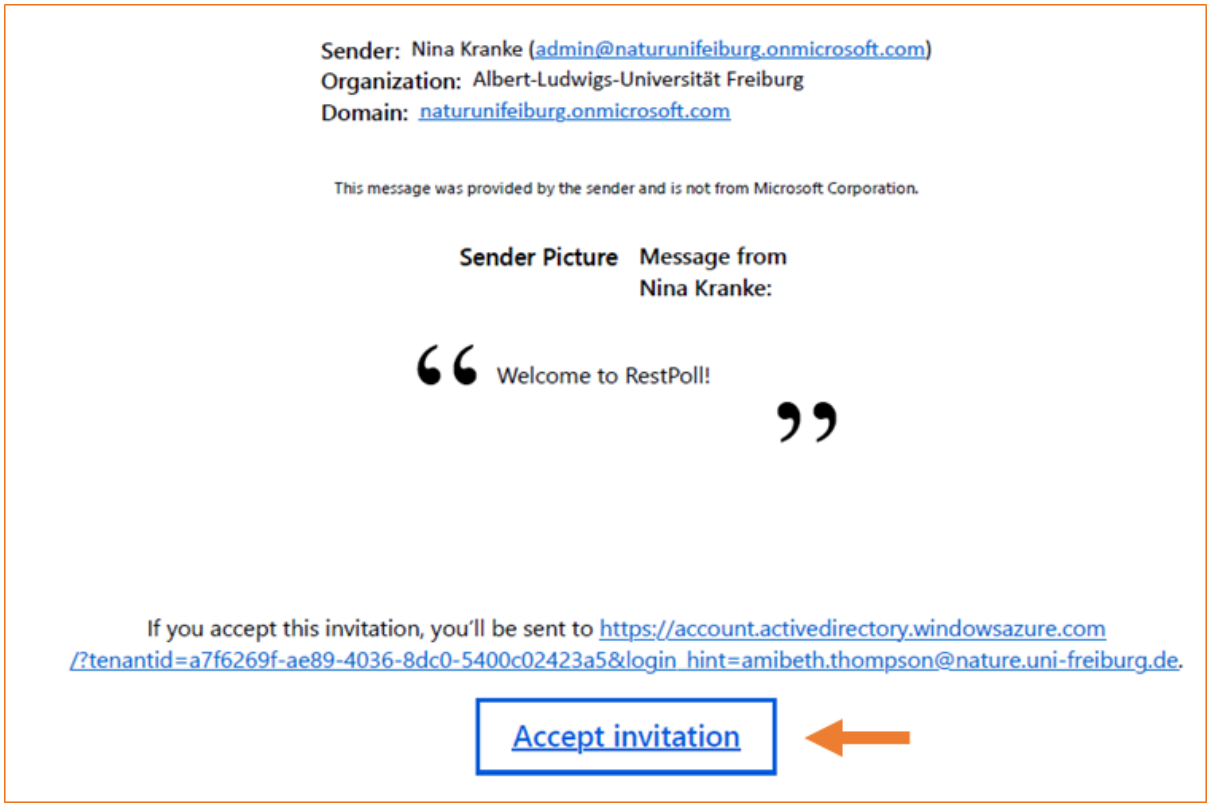
17. Accessing the RestPoll Drive

All file managers and other relevant RestPoll members will be invited to join the RestPoll Drive via an Email invitation. Please inform Amibeth Thompson (amibeth.thompson@nature.uni-freiburg.de) for access to the drive.

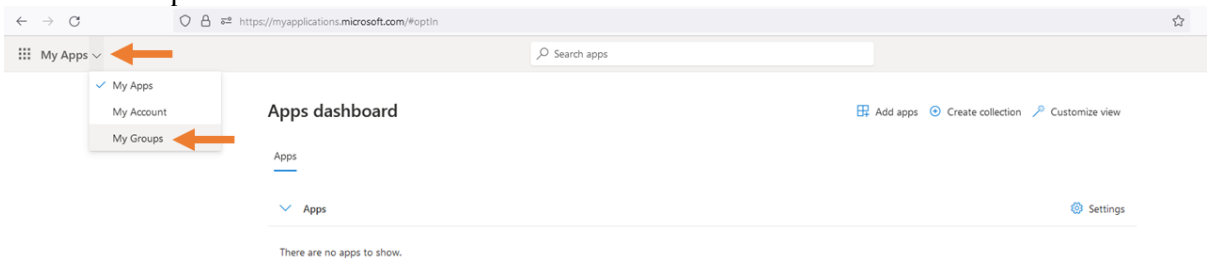
1. The email invitation will be from “Nina Kranke”. Please click “Accept Invite” and follow the instructions to sign in.

Funded by the European Union

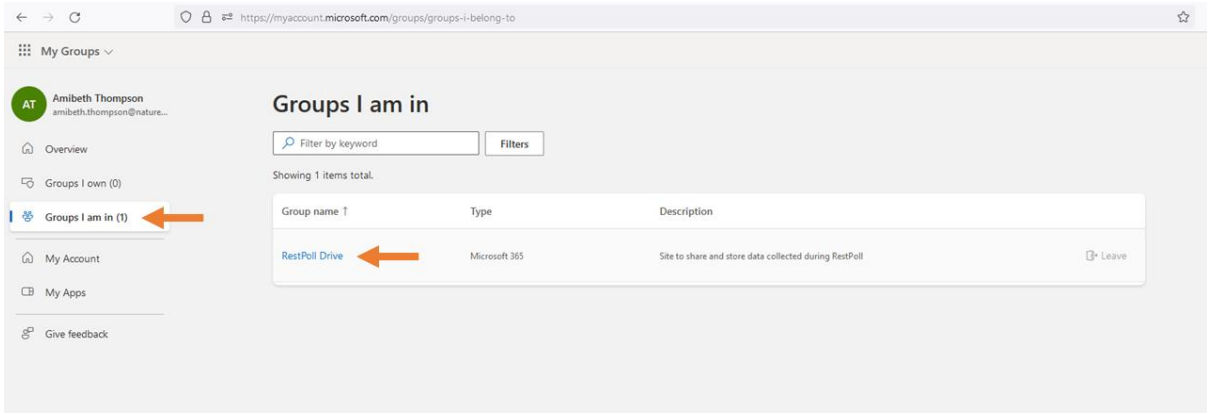
This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.



2. Once you are logged in you will see an empty App-Dashboard, i.e. without any apps. If you click on the "v" symbol in the upper left-hand corner, then on 'My Groups'.



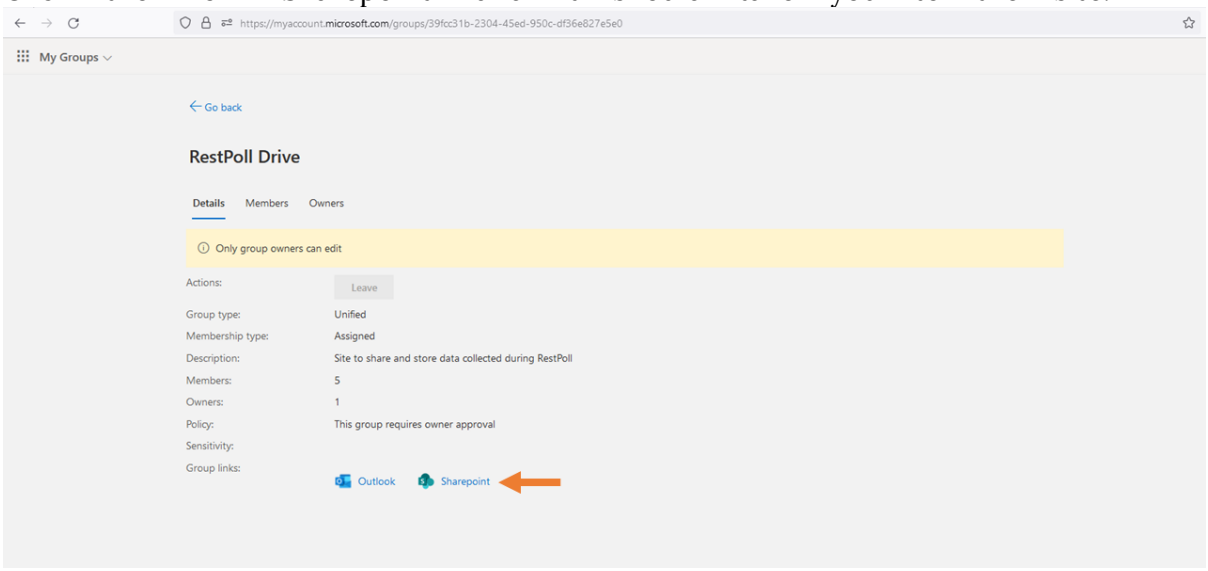
3. Afterwards, click on "Groups I am in (1)", where you should have access to "RestPoll Drive."



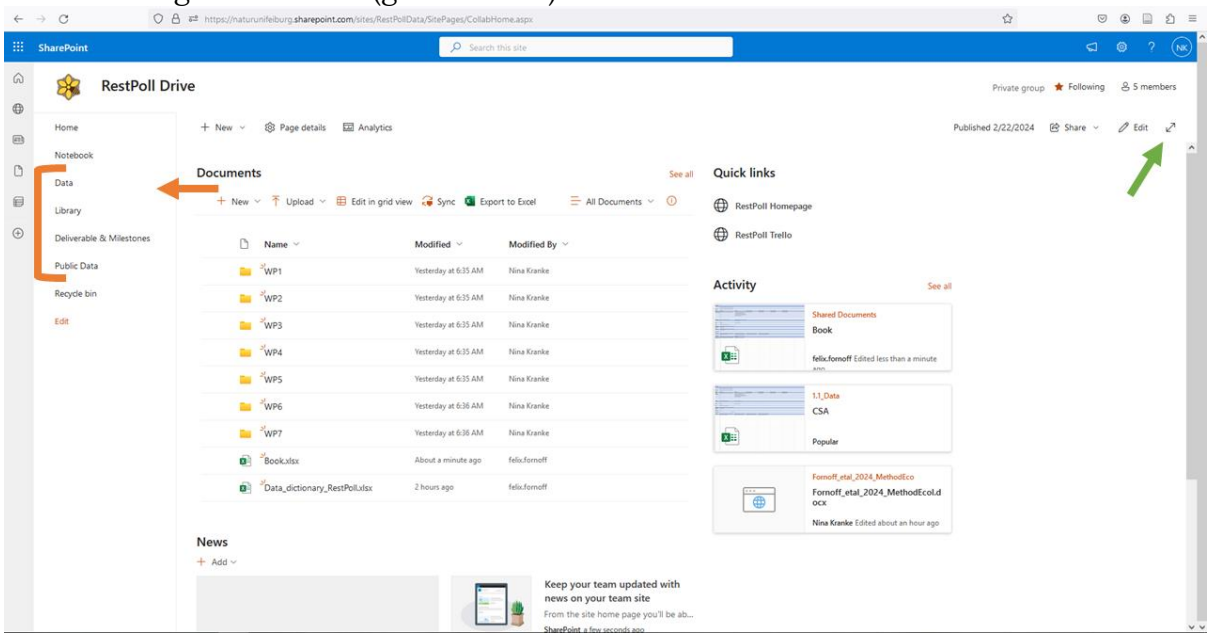
Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.

4. Click then on “Sharepoint” and it should take you to the site.



5. This will bring you to the RestPoll Drive homepage. The different library should be visible on the lefthand side of the page (orange arrow). If not, click the arrows on the right-hand side (green arrow).

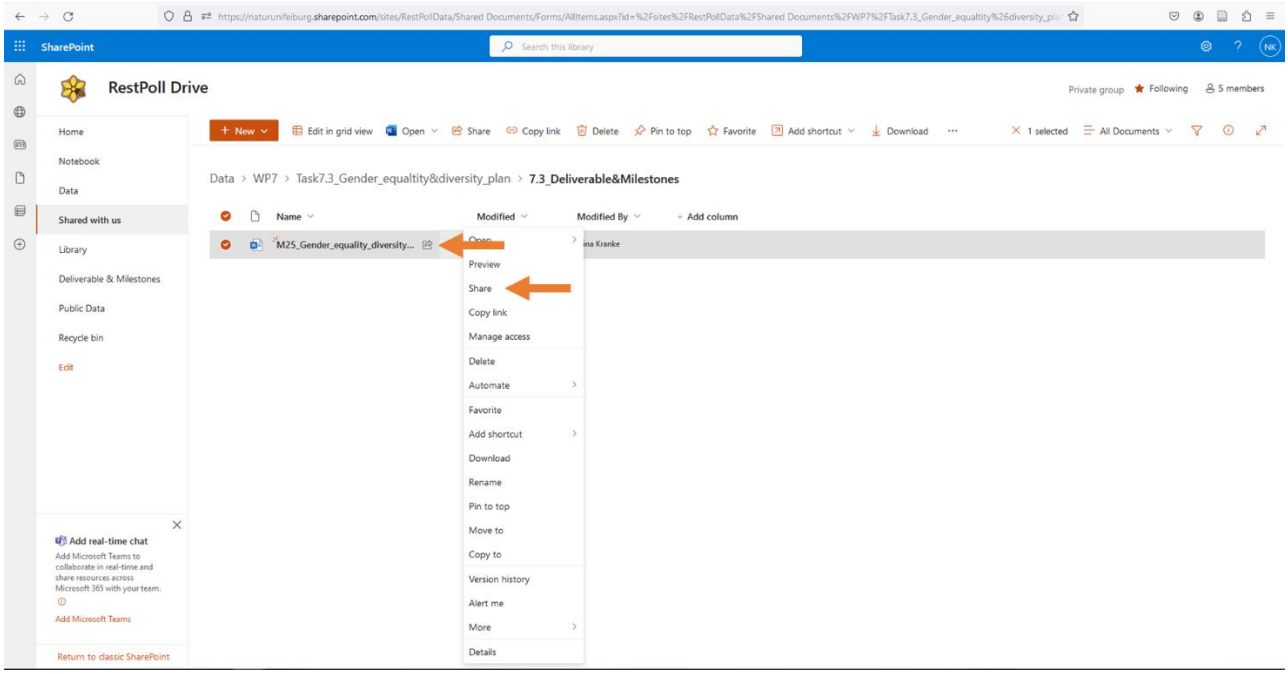


17.1. DATA SHARING

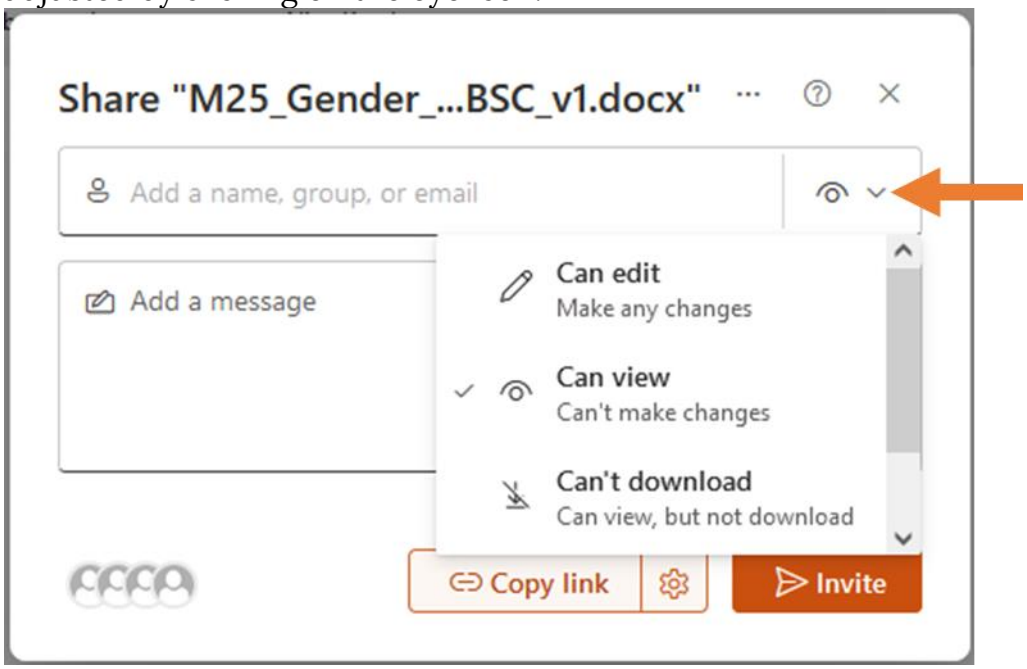
Only file managers will have full access to the RestPoll Data drive. However, file managers can share data (folders, files, etc.) with other members. To share data, click either the share icon or the three dots and “share”.

Funded by the European Union

This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.



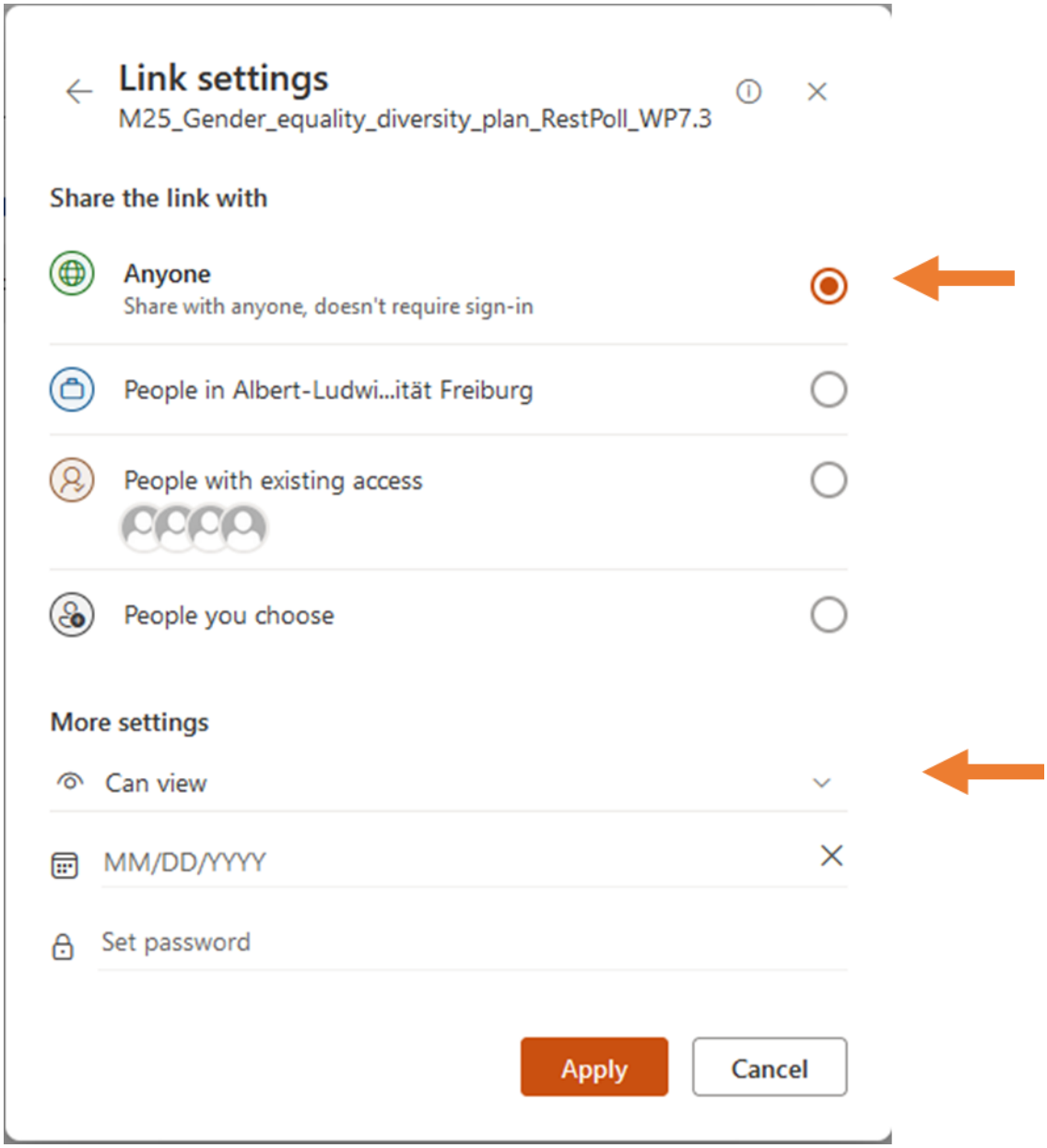
The standard settings are to share the date with “view” settings, but these can be adjusted by clicking on the eye icon.



An email can be sent or a link can be created and shared with anyone. A password or expiration date can be. More information can be found here: https://support.microsoft.com/en-us/office/share-sharepoint-files-or-folders-1fe37332-0f9a-4719-970e-d2578da4941c

Funded by the European Union

This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.



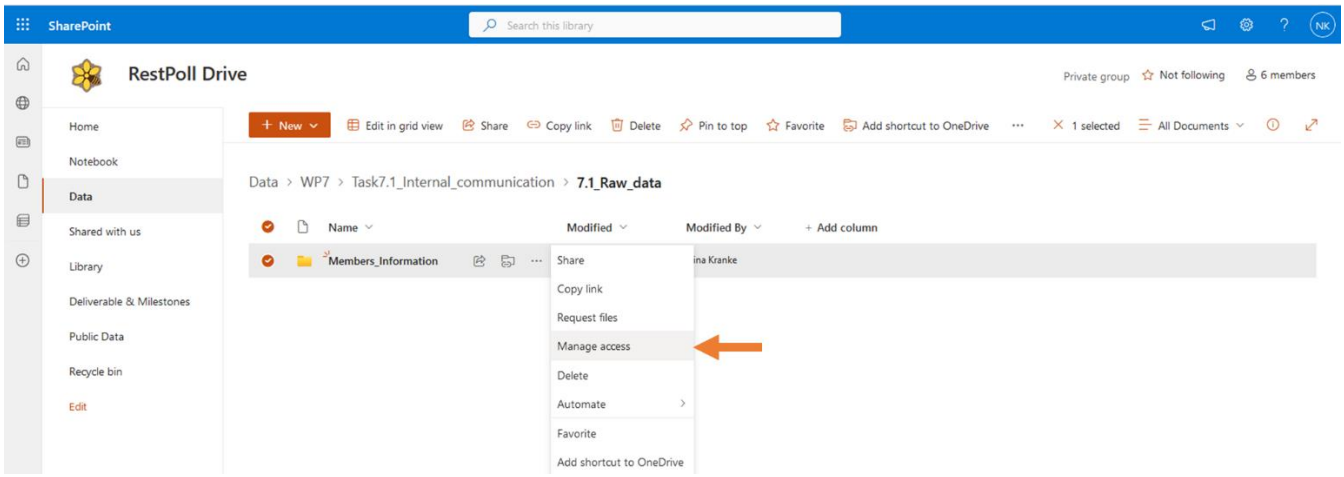
17.2. DATA SECURITY

Because all file managers have access to the RestPoll Drive, some data may need to have restricted access, such as data with sensitive information. Private folders can be created within the WP or Task folders with access only for specified members.

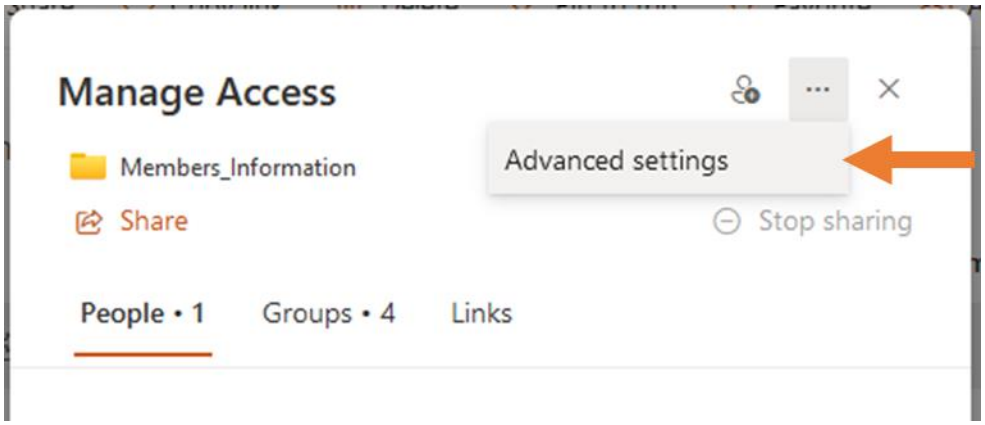
1. Create a folder where you would like to store your secure data. Click on the three dots (...) and click on “Manage access”.

Funded by the European Union

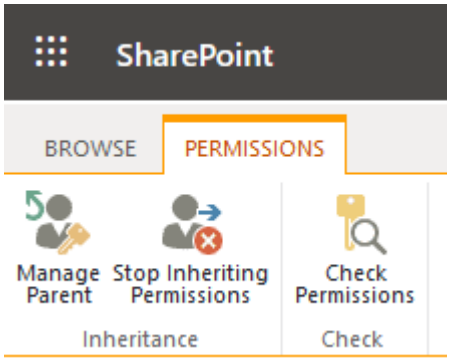
This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.

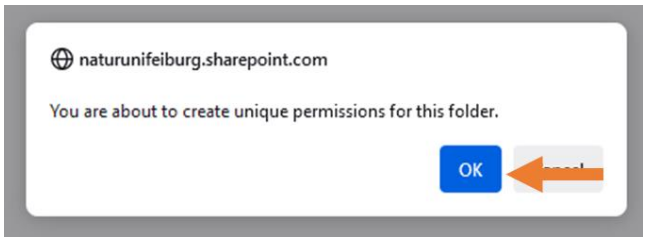


2. In the pop-up box, click again on the three dots (...) and then “Advanced settings”.



3. In the new window, click “Stop Inheriting Permissions”. In the pop-up window, click “OK”.

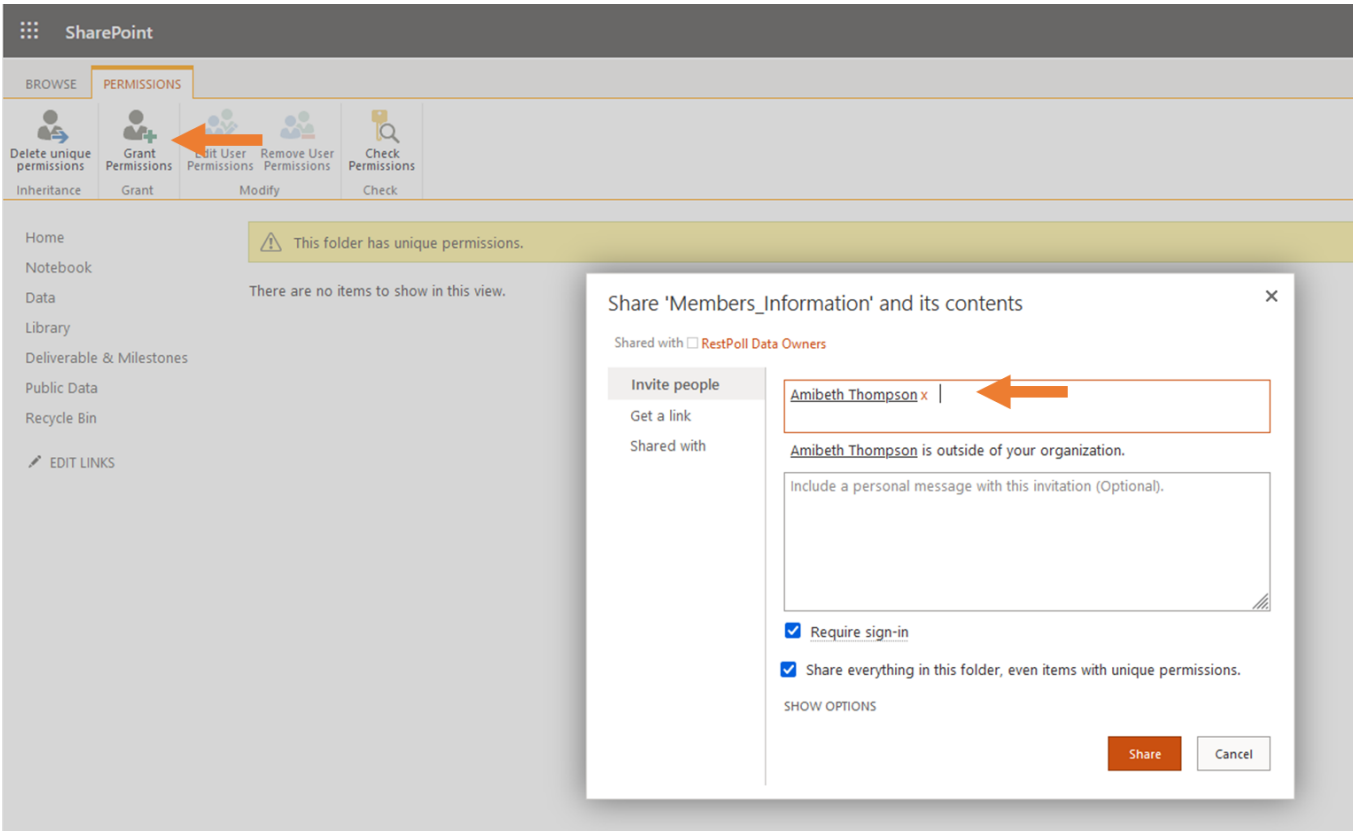




4. Select all current users and then click “Remove User Permissions”.
5. Once all users have been removed, click on “Grant Permission.” In the pop-up window, add the email address(es) of member(s) who should have access to the folder. Un-check “Require sign-in” and then click, “Share”.

Funded by the European Union

This project receives funding from the European Union’s Horizon Europe Framework Programme under project No. 101082102.



Funded by the European Union

This project receives funding from the European Union's Horizon Europe Framework Programme under project No. 101082102.