Root2Res

Root phenotyping and genetic improvement for rotational crops resilient to environmental change

Root2Res Data Repository

Jean-Pierre Cohan and Pierre Rochepeau (ARVALIS)

Timothy George and Fanny Tran (JHI)

Funded by the European Union

UK Research and Innovation

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Root2Res has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101060124. Its work is supported by Innovate UK through the Horizon Europe Guarantee scheme Grant Agreement No. 101060124 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) grant No. 23.00050.

Root2Res Data Repository provides all project's partners (and agreed licensed partners) a secured database specifically designed to accommodate and manage all Root2Res data, methods and information gathered from project activities. The storage procedures and repository structure are designed to ensure that datasets comply with FAIR principles and GDPR of the EU (and equivalent for the associated partners).

Deliverable Number	**Work Package / Task**
D8.4	WP8 / T8.4
Lead Beneficiary	**Deliverable Author (S)**
ARVALIS	Jean-Pierre Cohan
Beneficiaries	**Deliverable Co-Author (S)**
ARVALIS and JHI	Pierre Rochepeau (ARVALIS) Fanny Tran (JHI) Timothy George (JHI)
Planned Delivery Date	**Actual Delivery Date**
28/02/2023	28/02/2023

Type of deliverable	R	Document, report (excluding periodic and final reports)	
	DEM	Demonstrator, pilot, prototype, plan designs	
	DEC	Websites, patents filing, press & media actions, videos, etc.	
	DATA	Data sets, microdata, etc.	X
	OTHER	Software, technical diagram, algorithms, models, etc.	

Dissemination level	PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page))	
	SEN	Sensitive, limited under the conditions of the Grant Agreement	X

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, UK Research and Innovation (UKRI), European Research Executive Agency (REA) or Swiss State Secretariat for Education, Research and Innovation (SERI). Neither the European Union nor any other granting authority can be held responsible for them.

INDEX

FIGURES

1. Introduction

The Root2Res Data Repository aims to provide all project's partners (and agreed licensed partners) a secured database specifically designed to accommodate and manage all Root2Res data, methods and information gathered from project activities. The storage procedures and repository structure will be designed to ensure that datasets comply with FAIR principles (Findable, Accessible, Interoperable and Reusable) and GDPR of the EU (and equivalent for associated partners). The Data Repository is strongly linked to the Data Management Plan (DMP), which first version is due at the same time (M6). The DMP will be updated twice during the project lifetime (M36 and M58). On this occasion, some updates of the Data Repository may also be included.

2. Characteristics of the Data Repository

2.1. Location and access

To facilitate a rapid access to the Data Repository while complying with the FAIR and GDPR principles, its first version has been created in the project workspace using the "Microsoft SharePoint" technology (see Deliverable 8.1 and Figure 1). All partners have access to the workspace and the technology allows the workspace manager (Pierre Rochepeau; p.rochepeau@arvalis.fr) and the data manager (Root2Res.data.officer@arvalis.fr) to grant selective access to specific sub-sections and files if needed (before finalization of data files for example or depending on the rules set in the consortium agreement).

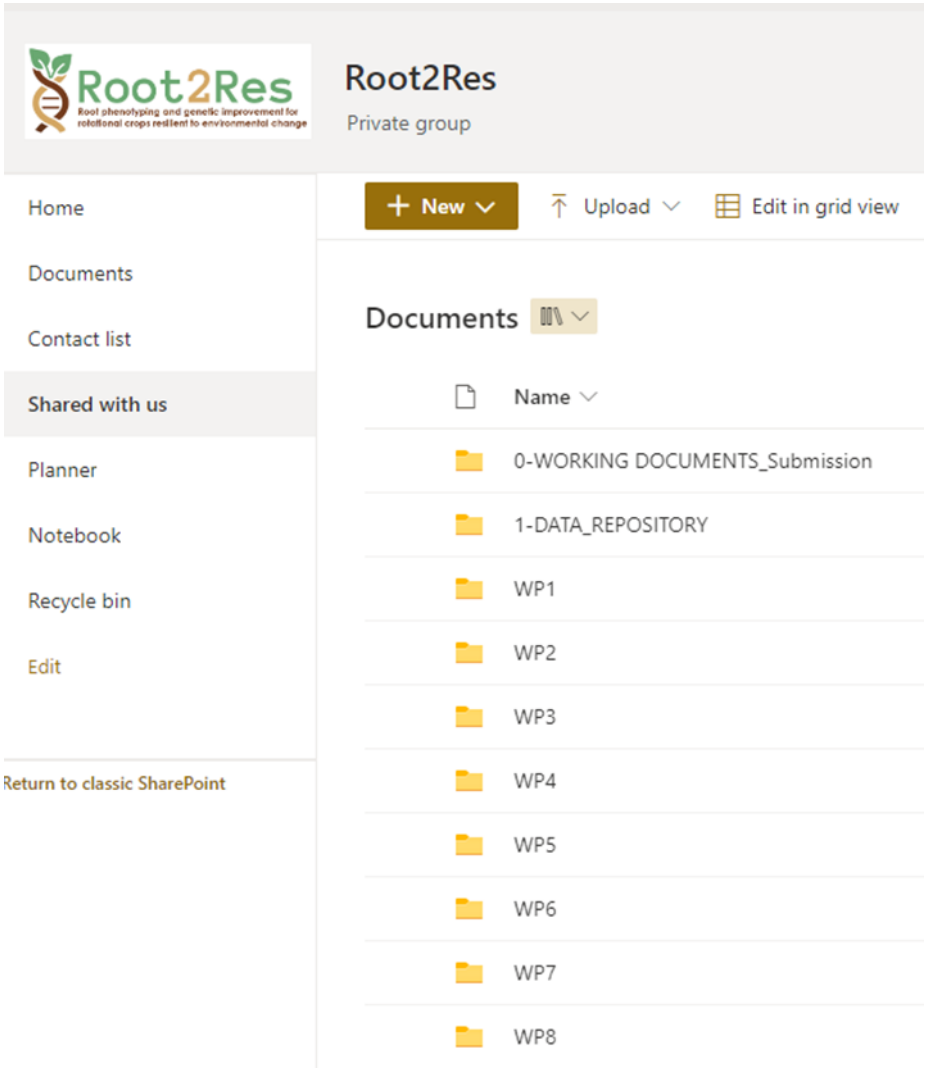

Figure 1. Location of the Data Repository in the general structure of the Root2Res workspace.

2.2. Internal structure

The internal structure of the Data Repository (Figure 2) has been designed to correspond to all the categories of data (Data Group) identified in the Data Management Plan (see Deliverable 8.3). These include:

- Raw Data (RA)
- Imaging Data (IM)
- Field and Phenotyping Data (FP)
- Omics Data (OM)
- Modelling Data (MO)
- Exploitation (EX)
- Communication and Dissemination (CD)

Additionally, a specific folder has been created to store all data templates included in the Data Management Plan (for datasets and metadata). The data register is also accessible at the root of the structure with specific access (see. 2.3).

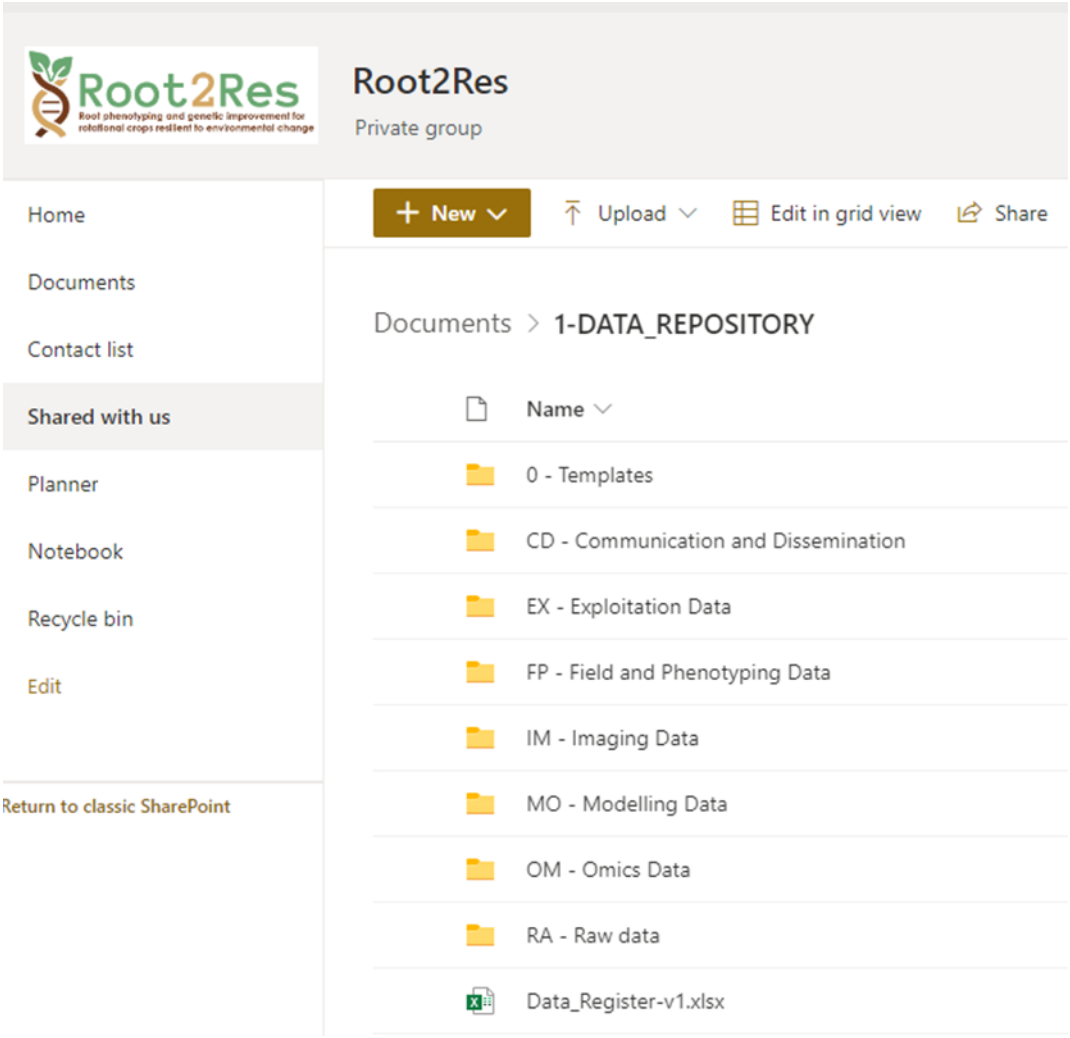

Figure 2. Internal structure of the Data Repository.

2.3. Compliance with the FAIR Principles and GDPR

FAIR principles (Findable, Accessible, Interoperable and Reusable) have been respected with the following functionalities of the Data Repository:

- **Findable:** The Data Repository is located in the workspace of the project, which is now the central tool to manage the Root2Res project for all partners.

- **Accessible:** since M3 (see Deliverable 8.1), all partners have access to the Root2Res workspace and, therefore, to the Data Repository. The Microsoft SharePoint technology allows the workspace manager (Pierre Rochepeau; p.rochepeau@arvalis.fr) and the data manager

(Root2Res.data.officer@arvalis.fr) to grant selective access to specific sub-sections and files if needed. The status of each file will be determined at its creation and will comply with the rules set up in the consortium agreement.

- **Interoperable:** all the data stored in the Data Repository will use the tools developed in the Data Management Plan (data and metadata templates, common set of variables using the ICASA standard completed with specific variables created for the project). At this stage of the project, the need for interoperability is limited to the crossed use of data files inside each category. The proposed data storage structure, divided in sub-sections for each data category, is suitable for this purpose (query on or merging of individuals files with the same format in one sub-section). When the need for crossed queries between sub-sections arises, an evolution of the data repository will be proposed (see Section 4).

- **Reusable:** The reusability of the data stored in the data repository will be guaranteed by 1) the use of data and metadata templates proposed in the Data Management Plan and 2) the use of the Data Register. This essential tool to make the Data Repository fully operational will be managed by the data manager (Root2Res.data.officer@arvalis.fr) and as outlined in the Data Management Plan. It will be accessible, as read-only, to all partners and will include (Figure 3):

    - a unique identification code of the dataset;
    - its dates of creation and validation;
    - the list of partners involved in its creation;
    - its status of availability after the end of the project;
    - a brief description.

DATA REGISTER

Dataset-ID	Date of creation	Date of validation	Partners involded	Availability	brief description
Example					
FP-WP2-T1-00001	30/06/2023	15/09/2023	ARVALIS	Open access after publication/exploitation	Fied dataset acquired in ARVALIS experiment n°1 (Phenofield site) for WP2 task 1 (test of phenotyping tools)

Figure 3. Extract of the Data Register.

As all data management systems within the Root2Res project, the Data Repository will comply with the obligations of the EC GDPR and its equivalent for the associated partners. This will be managed by the data manager (Root2Res.data.officer@arvalis.fr).

3. Link with other deliverables

3.1. Data Management Plan (DMP)

As a central tool to manage data, the Root2Res Data Repository is included in the Data Management Plan (Deliverable 8.3). Upon updates of the DMP, some updates of the Data Repository may also be necessary.

3.2. Dissemination and Exploitation dissemination plans

Dissemination and exploitation materials are described in the Dissemination and Exploitation Plans (Deliverable 7.2). They have been identified as data categories in the Data Management Plan and so, will be properly stored in the Data Repository, with the appropriate accessibility.

4. Further updates

This first version of the Root2Res Data Repository fulfils the requirements of the partners for at least the first year of project implementation. Indeed, the current needs for data interoperability are limited to dataset categories individually. Nevertheless, it is forecasted that the needs for crossed uses of data from different categories will emerge at some stages of the project, such as, the complex collation of field, omics, and modelling data to produce the genetic toolbox. This will require the Data Repository to evolve possible using alternative technology (e.g., PostgreSQL), which would allow complex and elaborated crossed data queries. Such evolution requires deep conversations between WP and crop leaders to define the exact functionalities required. These discussions will be conducted during the year 2023. It is important to note that the structure option outlined in the current Data Repository (i.e., unique ID, Data Register, structure in sub-section) will facilitate the transition into the new technology, whenever required.