Deliverable			**D8.1**
Deliverable title: Data Management Plan			
Task	**WP8 – T8.1 dissemination** **WP8 – T8.2 exploitation**		
Task Leader	**IAAC and Sony**	Planned Date	**30 April 2018**
		Effective Date	**21 June 2018**

	Written by	**Reviewed and approved by**		**Authorized by**
Name / Surname	**Peter Hanappe (Sony) with the support of Olivier Degrand (FEI)**	All the consortium		Jonathan Michin (IAAC)

1. Overview

Most of the data produced in the ROMI project fall into one of the following five categories:

1. **Design files**: This category covers any of the following: the documents describing the hardware & software design, the experimental set-up, the protocols of the field studies, the hardware design etc.

2. **Source code**: The program files, the analysis scripts...

3. **Collected & created data**: This includes the data collected in the indoor & outdoor field studies and in the phenotyping experiments, but also the collection of generated images used in the machine learning (see below). It is important also to distinguish between:
   - **Experimental data sets**: These sets include data that is collected during development, testing, and exploring. The data may be voluminous, noisy, unstructured, and incomplete.
   - **Systematic data sets**: This data has clear scientific value. These data sets were systematically collected, cleaned, and prepared and may therefore have more value for the scientific community. The data sets that are used in publications by the consortium fall into this category.

   The (collected and created) datasets that are most useful for publication are:

   1. **Outdoor Studies Data**: Data collected using the NERO drone and LettuceThink robot during the outdoor field studies: the photos taken automatically (drone, robot and time lapse cameras) and the environmental data.
   2. **Plant Models Images**: The synthetic images created to train deep learning networks or to test computer vision algorithms.
   3. **Scanning Data**: Data collected using the plant scanner: Images captured for selected plants in indoor/outdoor settings using various 2D and 3D cameras images to evaluate the algorithms of the scanner.
   4. **Adaptive Systems Data**: Data collected during the experiments of the adaptive systems (development, testing, reference).
   5. **Phenotyping Data**: Data collected through the manual scoring from laboratory experiments on Arabidopsis thaliana (and possibly other plants).
   6. **Plant Growth Data**: The datasets used to analyse the plant growth. This dataset will probably be a subset of the other datasets (1, 2 & 4 above).

4. **Analysis results**: This category covers the manuscripts that document the analysis and evaluation of the data. This may include figures, selected datasets and source code that are used to illustrate the texts.

5. **Media files**: This contains photos, videos, and audio recordings that document the ROMI project and visual material for presentations. These data are not already classified in the other categories

These datasets may overlap. This is most obvious for the Plant Growth Data that uses data from the field studies but an overlap may also exist, for example, between the Scanning Data and the Outdoor Studies Data.

2. File formats and associated software

The table below displays the main data types that will be used, their purpose, and the compatible software.

File format	Data type	Software	Purpose	More info
JPEG	Bitmap images	An image viewer (ex. Gimp)	Data exchange of compressed images	
TIFF	Bitmap images	An image viewer (ex. Gimp)	Data exchange of lossless, compressed and layered images	
MPEG	Videos	A video player (e.g. VLC)	Data exchange of video recordings and image sequences	
MP3	Audio	Any audio player	Interview recordings	
PCD	3D data (point clouds)	PCD viewing application (e.g. pcl_viewer or CloudCompare).	Data exchange of point cloud data	http://pointclouds.org/documentation/tutorials/pcd_file_format.php
PLY	3D data (triangle mesh)	3D visualisation program (ex. ParaView).	Data exchange mesh data	https://en.wikipedia.org/wiki/PLY_(file_format)
JSON	Data-interchange format	A text editor	Data exchange of meta-data, measurements and parameter files.	https://www.json.org/
CSV	Measurements	A text editor or spreadsheet program (e.g. OpenOffice)	Data exchange of measurements.	https://en.wikipedia.org/wiki/Comma-separated_values
PDF	Office documents	A PDF Reader, including Evince (Linux), Acrobat Reader (MacOS &	Publication of design documents, manuals, annotations, research results, etc.	https://en.wikipedia.org/wiki/Portable_Document_Format

		Windows) and many more.		
L-Py	Source code for L-Py.	Text editor and the L-Py simulation software.	Plant modeling code: L-systems constructions mixed with the Python high-level modeling language.	http://lpy.readthedocs.io/en/latest/
Python	Source code for the Python interpreter	A text editor and the Python interpreter (see also "Software dependencies" below).	Software development, data processing algorithms, and parameter files.	https://www.python.org/
C/C++	C/C++ source code	A text editor and C/C++ compiler (e.g. GNU Compiler Collection).	Software development and data processing algorithms.	https://gcc.gnu.org/
R-code	Source code for the R interpreter.	A text editor and the R environment.	Data exploration, plotting and statistical analysis.	https://www.r-project.org/
Javascript	Javascript source code, for	Text editor and Javascript interpreter (web browser or NodeJS)	Software development, in particular web-applications (client-side)	https://developer.mozilla.org/bm/docs/Web/JavaScript

Notes:

- **Open Source:** Open Source software exists to view the all of the datasets and file formats.
- **Documentation software**: All the software that are mentioned above come with their own user manuals. No additional documentation is needed.
- **Software dependencies**: The Python, C/C++ and Javascript source code may depend on additional software libraries. The documentation of each package developed by ROMI will list the software requirements. We will aim to limit these dependencies to widely available, standard and Open Source libraries.
- **Portability**: Python code will be platform independent. All the developed C/C++ software will run at least on the Linux operating system and care will be taken to make the code as portable as possible.
- **Alternative file formats**: Some 2D and 3D data may also be provided in common alternative file formats including PNG (2D images), STL and OBJ (3D mesh data). Converters to transform between these formats and the formats listed above are available.
- **Office documents**: All office documents will be provided in the PDF file format. In addition, some files may be provided in an editable format, using one or more of the formats below:

File format	Data type	Software	More info
OpenDocument	Office documents	LibreOffice, OpenOffice	https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office
Office Open XML	Office documents	Microsoft Office Suite, LibreOffice, OpenOffice	https://en.wikipedia.org/wiki/Office_Open_XML
Google Docs, Sheets, and Slides	Office documents	Google's online office suite	http://docs.google.com/

3. Main storage locations

By default, the data that is produced by the ROMI project will be stored in the consortium's Google Drive and the code that is produced will be stored in the project's Github repository. The table shows the additional storage locations that will be used.

Data	Location	Notes	Duration
Source code	Github	For public source code. Address https://github.com/romi/	10 years
Collected data	Google Drive	For public datasets.	10 years
Documents	Google Drive	The PDF versions of the public documents.	10 years
Collected data	Standard repositories	Datasets that are used in publications or that have a clear interest to the wider scientific community will be made available through standard data repositories.	Unlimited (defined by repository).
Collected data	Project server	For large datasets, the Google Drive may be too expensive. A project server will be set-up.	5 years
Documents	GitHub	Some design documents and user manuals may be stored with the software in the Github repository.	10 years
Source code	Local code repository	For unpublished code (e.g. work in progress)	During the project

Collected data	Local storage	Data will be stored locally by the partners for data processing and for backup.	During the project
Media files	Google Drive	For public files.	10 years

Notes:

- **Alternative retrieval method**s: All datasets will be made available through standard web-based interfaces (using http-based URLs). In addition, some datasets may be made available by addressing the API of the storage back-ends used by the consortium. This API may give more powerful search and indexing functionalities. However, it might be necessary to request access to the consortium in order to create an account on the back-end server before using this functionality.
- **Alternative code repositories**: Some partners may use additional code repositories for their existing software or legacy code. In that case, we will clearly document where and how to retrieve this code, for example by using repository redirections.

4. File organisation

The organisation below has been defined by SONY at the beginning of the project. It will be confirmed during the M12 meeting by the whole consortium.

For the sake of starting with a initial scheme, we propose the following naming strategies and directory organisation.

Filenames
Below are two examples of filenames that should be self-explanatory:

romi-wp2-lettucethink-manual-20180426.pdf

romi-wp7-fieldexperiments-chatelain-fieldlayout-20180426-1.pdf

More formally, the naming scheme is:

romi-<workpackage>-<modulename>[-<location>]-<description>-<date>[-<version|index>].<filetype>

With:

- *workpackage*: wp1, wp2, …
- *modulename*: example: plantmodeling, adativesystems, nero, lettucethink, fieldstudies, …
- *location*: When relevant, mostly for data gathered in field studies: chatelain, iaac, cnrs, …
- *description*: One word to describe to contents: measurements, fieldlayout, …
- *date*: year, month, day
- *version* or *index*: In case of multiple files or versions.
- *filetype*: Standard 3 character file extension.

Folder structure
The files are organised in a folder structure that reflects the work packages, the modules, and the type of file. For example:

WP2/LettuceThink/Design/

WP2/LettuceThink/Media/Photos

WP2/LettuceThink/Media/Videos

WP7/Chatelain/Design/

WP7/Chatelain/Data/

WP7/Chatelain/Results/

WP2/Chatelain/Media/Photos

WP2/Chatelain/Media/Videos

The data files will require more levels of organisation. They will be documented in future updates of this document.

5. Accessibility and licensing

⇨ **Please refer to the associated Excel file – see the Annex**

6. FAIR data

Making data findable, including provisions for metadata:

- ⇨ *Outline the discoverability of data (metadata provision)*
- ⇨ *Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?*
- ⇨ *Outline naming conventions used*
- ⇨ *Outline the approach towards search keyword*
- ⇨ *Outline the approach for clear versioning*
- ⇨ *Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how*

	Design files	**Source code**	**Collected & created data**	**Analysis results**	**Media files**
Discoverability	Description on ROMI web site + web search engines	Description on ROMI web site + web search engines + GitHub search	Experimental data sets: Description on ROMI web site + web search engines Systematic data sets: Description on ROMI web site + web search engines + search	Description on ROMI web site + web search engines	Description on ROMI web site + web search engines + paper dissemination tools (booklet, posters, flyers …)

Identifiability	Web URLs	Web URLs	Experimental data sets: Web URLs Systematic data sets: Web URLs and DOIs provided by data repositories	Web URLs	Web URLs
Naming convention	File names include module name and version or date.	Project name includes the module name.	Data files will be structured in hierarchical folders with clear names and dates		File names
Keyword search	Description + web search engines				web search engines
Versioning	Date in file name	Git + standard versioning schemes	Date in file name	Date in file name	Date in file name
Metadata	Keywords on web site	Keywords on GitHub	Experimental data sets: High level description + keyword + description of formats on web site Systematic data sets: relevant metadata standard will be used	The relevant metadata standard will be used	Keywords on web site

Design files

To our knowledge there is currently no standard platform to make design files and manuals available. The consortium will take the following steps to make these files findable:

- The design files will be accessible from the ROMI web server with a clear description, file name and meta-data where possible.
- In addition, the consortium partners may maintain their own set of web pages to document separate hardware and software components of the project (NERO drone, L-Py software, LettuceThink robot…). The relevant design files will also be accessible from these pages with clear descriptions.

- During the project, the partners will continue to investigate the available documentation platforms, such as Readme.io, GitHub pages, and other, to make the design files available through industry and community-supported solutions.

We believe that with the strategy above, the design files should be easily findable using web search engines.

Source code

The source code will be made available through industry standard repositories such as GitHub. The repository will be documented with clear descriptions and keywords which makes the code easily discoverable. Documentation on how to install and use the software will be part of the repository. Standard versioning practices will be used to keep trace of the stable versions of the code.

Collected & created data

We intend to share our dataset on the publicly accessible servers combined with descriptive metadata. Because main objective of the dataset is the evaluation of the algorithm during the development phase, no special efforts will be made to associate DOIs for identification and citability of the dataset. However, the data will be clearly accessible and easily findable from the project's web site (Google Drive). Folders and files will be organized in a hierarchical and clear structure. Files will be uniquely identifiable and versioned by using a name convention consisting of project name, dataset name, method used, ID, place and date.

For non-experimental data sets we will upload the data to field specific repositories and use the recommended metadata. Existing repositories can be found, for example, through Registry of Research de Repositories (Re3Data, Re3data.org). A relevant example for the project is TAIR - The Arabidopsis Information Resource (www.arabidopsis.org)

Analysis results

All results will be introduced in project deliverables. They will be uploaded in the project Google Drive.

For published result we will upload the data to field specific repositories and use the recommended metadata.

Media files

All data (photo, sounds, video …) used for illustrating the project will be available on the project website. They will also could be accessible in all sites which will introduce the pro

7. Making data openly accessible

- ⇨ *Specify which data will be made openly available? If some data is kept closed provide rationale for doing so*
- ⇨ *Specify how the data will be made available*
- ⇨ *Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?*
- ⇨ *Specify where the data and associated metadata, documentation and code are deposited*
- ⇨ *Specify how access will be provided in case there are any restrictions*

	Design files	**Source code**	**Collected & created data**	**Analysis results**	

Openly available	See associated Excel file			
How made available	See Section 3			
Needed tools & documentation	See associated Excel file and Section 2			
Location	See associated Excel file and Section 3			
Access in case of restrictions	The files will be made available when the design is finalized	Intermediate versions of source code may not be public until the next stable release. External collaborators may request access by email.	Data sets and analysis results may be made available after an embargo period to allow for scientific publications.	No restriction

8. Making data interoperable

⇨ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

⇨ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

	Design files	**Source code**	**Collected & created data**	**Analysis results**	**Media files**
Interoperability	NA	NA	Experimental data sets: widely used file format will be used For systematic data sets, additional ontologies may be used to facilitate the access with field-specific software tools	Ontologies may be used to facilitate the access with field-specific software tools	NA
Ontologies	NA	NA	To be defined. See for example planteome.org	To be defined. See for example planteome.org	NA

Source code

The software will be made to run at least on Linux, an industry-standard and free operating system. Efforts will be made to ensure that all the source code is portable to other platforms, in particular MacOS and Windows.

Collected & created data

The data, metadata and documentation are compliant to disciplinary standards and open file formats for easy interoperability and re-use.

9. Increase data re-use (through clarifying licenses)

- ⇨ Specify how the data will be licenced to permit the widest reuse possible
- ⇨ Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- ⇨ Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- ⇨ Describe data quality assurance processes
- ⇨ Specify the length of time for which the data will remain re-usable

The project data will be made available between 5 and 10 years after the end of the project.

	Design files	**Source code**	**Collected & created data**	**Analysis results**	**Media files**
Licensing	Creative Commons license	Open Source license	Creative Commons and/or Open Data license	Creative Commons and/or Open Data license	Creative Commons license
Embargo			Available last three months of project end. Data sets and analysis results may be made available after an embargo period to allow for scientific publications.		NA
Useable by third-parties	Yes	Yes	Yes	Yes	Yes
Data quality assurance process	Best effort, no guarantees	Best effort, no guarantees	Experimental data sets: Internal review Systematic data sets: Following best practices of targeted scientific community	Following best practices of targeted scientific community	NA
Length of time	5 or 10 years	10 years	5 years (large, experimental data sets), 10 years (other)	10 years	10 years

Design files: The design files will be made available as under a Creative Commons license. The files will be made available during the last three months of the project, or before according to the needs to collaborate with third parties. A snapshot of the files will be kept on the ROMI project server for at least 5 years.

Source codes

The software will be made available under a Free or Open Source license. It will remain accessible on cloud-based code repositories and is freely useable by third parties under the conditions laid out in the license. The source code will be copied to the public repository only after significant testing by the project partners, and the code will have well-defined versions to aid with the tracking of problems. The partners will make the best effort in assuring the quality of the code but cannot guarantee usability and do not give any warranty against failures, data loss or damages that may have been caused by the software.

Collected & created data

The data will be licensed under Creative Commons CC BY 4.0 International. All data will become available during the last three months of project end. Parts of the data can become available even before due to journal publications. An embargo period to allow for scientific publications may be applied.

The data can be re-used by other scientists. Neighboring disciplines and interdisciplinary research groups might also be interested, because of our chosen methodological approach.

The data quality is ensured by different measures. These include validation of the sample, replication, comparison with results of similar studies and control of systematic distortion.

As open formats are used for data archiving, the data will remain re-usable until the repository withdraws the data or goes out of business.

Analysis results

The process of validation of a project deliverable is accurately described in the consortium agreement. All deliverables are public.

Media files

The media files will be licensed under the Creative Commons CC BY 4.0 International license. The files will become available on the web site during the course of the project and remain available as long as the web site is online. There is no specific embargo period but before the files are uploaded to the server we will assure that those concerned by the files (ex. because they appear in a photo) have given their agreement for its distribution.

10. Allocation of resources

- ⇨ Explain the allocation of resources, addressing the following issues:
- ⇨ Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- ⇨ Clearly identify responsibilities for data management in your project
- ⇨ Describe costs and potential value of long term preservation

Each partner is responsible for uploading data in evoked above repository. SONY and IAAC will monitor the progress state of uploads before each General Assembly meeting.

The partner FEI will take in charge during the project the hosting costs related to Google Drive. After the project, it will the responsibility of both IAAC and Sony and afterwards of the start-up which will be created to take in charge these costs.

	Design files	**Source code**	**Collected & created data**	**Analysis results**	**Media files**
Estimated cost	Up to 3% of the project time.	No additional costs	No cost for making data FAIR. Preparation and cleaning of data sets take up to 10% of the project time	Preparation and documenting of result sets take up to 15% of the project time	Costs of a professional of images (if relevant)
Responsibilities	Each task leader is responsible for preparing the design documents, source code, data sets and analysis results				IAAC
Cost of long term preservation	Low (120 €/yr)	Free (GitHub)	High for large data sets (1200 to 10 000 €/yr)	Average (120 to 1200 €/yr)	No cost
Value of long term preservation	Average to high	High for production code. Debatable for explorative projects.	High for systematic data sets. Debatable for experimental data sets.	High (scientific research)	Weak (evolution of systems)

Design files

The preparation of the design files and of the associated manuals and explanatory documents is a very time-consuming process that can take up to an estimated 3% of the project time. The responsibility of preparing the documents for a given component falls onto the task-leader of that component. The work is an essential part of documenting the project (short-term and long-term) and is considered part of the description of work.

Source codes

Documenting and publishing the source code is part of the established work processes. It incurs no additional costs. The long-term storage will be mostly valuable to researchers developing to benchmark their algorithms against ours.

Collected & created data

The cleaning up and sorting of data sets can be costly in time but we estimate there are no costs for data preparation to be FAIR. Because we use the basic identification, annotation and storage scheme for these datasets, the procedure will be integrated in the daily work processes. Local team leaders at each site will be responsible for the data management plan updates. They will inform the Project Management Team for any anomalies or major changes. IT managers at the local sites will be responsible for backup and storage.

Long term preservation of small to medium sized data sets will result in no additional costs other than repository charges. For large data sets that are stored in Google Drive or on a project server, the costs may increase significantly. The preservation of these data sets on publicly accessible servers may be guaranteed only for five years after which the data will be backed up and stored by the relevant consortium partner and made available to third parties upon request.

The long-term storage will be mostly valuable to researchers developing to benchmark their algorithms against our data.

Analysis results

It is the responsibility of IAAC to maintain the project website after the project.

Media files

So as to prepare dissemination tools, the consortium could appeal a professional of images and sounds (if relevant) to prepare nice pictures, videos … Nevertheless, due to the constant evolution of systems which will be developed in the project, the value of long term preservation is quite weak.

11. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Design files

The design files will be stored on the project server that will be backed-up on a regular basis.

Source codes

Software will be stored a local repository during the development phase. Tested and versioned software will be copied to a cloud-based code repository over encrypted connections.

Collected & created data

During the project data will be saved on an institutional RAID-based server with backup to the institutional server or cloud-storage on a regular basis. The additional costs for storage and backup will be made to purchase a RAID-based network-attached storage disk.

We do not expect to collect any sensitive data. In case of doubt, the issue will be raised to the Steering Committee who will define the procedure to create an anonymized dataset. In such cases, access to sensitive data is granted only for project members with clearance through non-disclosure agreements. In addition, rightful access is ensured through secure passwords and encoding of the folders and files within cloud storage. Data transfer is secured via HTTPS protocol.

Analysis results

See the paragraph on Collected & created data.

Media files

See the paragraph on Design files.

12. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Design files

NA

Source codes

NA

Collected & created data

Informed consent for data sharing and long-term preservation is included during data collection. Sensitive data will be separated as soon as possible and kept secure.

Analysis results

The results, which are based on the collected and created data, will be anonymized following the guidelines in the Ethics deliverable.

Media files

NA

13. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

14. Annex – the Excel Data Management file

Code name	Title	Description	Purpose	Partner name	collected /created	Origin of the data	Category and function (software, data for evaluation, research data, manuscripts, dissemination material, administrative data	Usefull for publications	Profile of the potential users outside the consortium	Types (image, plug-in, video, doc, library, source code ...)	Format	Software needed to access the data	size	Privacy level (consortium, public, partner ...)	Storage - repository (at the level of partner)	Storage - repository (at the level of the consortium)	Storage - publicly accessible disciplinary repository Metadata (yes/no) DOI (yes/no)	Destroyed at the end of the project ?	Duration of preservation
LettuceThink																			
LT-DESIGN	LettuceThink Design	Hardware design of the LettuceThink robot, including the mechanical, electrical and electronic components.		SONY	Created	Project	Design files		End-users, designers	Text, pictures, figures	PDF	PDF reader	< 100 MB	Public	Local file server	Cloud storage (Google Drive and/or Github)	NA	No	10 yrs
LT-CODE	LettuceThink Code	Default software layer of the Lettuce Think robot, based on the Robotics		SONY	Created	Project	Source code		End-users, application	code	Python, C/C++	Text editors	< 100 MB	Public	Private code repository	Cloud-based code repository (Github)	NA	No	10 yrs
NERO drone																			
NERO-DESIGN	NERO Design	Hardware design of the NERO robot, including the mechanical, electrical and electronic components, and		IAAC	Created	Project	Design files		End-users, designers	Text, pictures, figures	PDF	PDF reader	< 100 MB	Public	Local file server	Cloud storage (Google Drive and/or Github)	NA	No	10 yrs
NERO-CODE	NERO Code	Software developed for NERO		IAAC	Created	Project	Source code		End-users, application	code	Python, C/C++	Text editor	< 100 MB	Public	Private code repository	Cloud-based code repository (Github)	NA	No	10 yrs
Specific algorithms for the robot and the drone (applications)																			
WEEDING-CODE	Weeding Algorithm	The source code of the weeding app for the LettuceThink robot		SONY & INRIA	Created	Project	Source code		End-users, application developers,	code	Python, C/C++	Text editor	< 10 MB	Public	Private code repository	Cloud-based code repository (Github)	NA	No	10 yrs
LT-EXPLORE-CODE	LettuceThink Adaptive Algorithms	Active vision system and models for exploration strategies for the LettuceThink robot		UBER & SONY	Created	Project	Source code		Application developers, researchers	code	Python	Text editor	< 100 MB	Public	Private code repository	Cloud-based code repository (Github)	NA	No	10 yrs
NERO-EXPLORE-CODE	NERO Adaptive Algorithms	Active vision system and models for exploration strategies for the NERO drone		UBER & IAAC	Created	Project	Source code		Application developers, researchers	code	Python	Text editor	< 100 MB	Public	Private code repository	Cloud-based code repository (Github)	NA	No	10 yrs
Outdoor field studies for usability and testing																			
FARM-DESIGN	Outdoor Studies Design	Design of the outdoor studies for LT & NERO: Site configuration, species and varieties, seeding/planting		Chatelain & IAAC	Created	Project	Design files		Researchers, farmers	Text, pictures, figures	PDF	PDF reader	< 100 MB	Public	Google Drive	Cloud storage (Google Drive and/or Github)	NA	No	10 yrs
FARM-OBSERVATIONS	Outdoor Studies Observations	Tables gathering evaluation data and comments	To have feedback on the usability of the tools ; to collect qualitative data on the state of the test	Chatelain & IAAC	Collected	Project	Collected and created data		Researchers, farmers	Tables and docs, pictures	PDF, JPG, MPEG	PDF reader, image viewer, video player	< 1 GB	Public, consortium	Google drive	Google Drive project server	NA	No	10 yrs
FARM-DATA	**Outdoor Studies Data**	**Data obtained using the NERO drone and LettuceThink robot during the outdoor field studies: Photos taken automatically (drone, robot and timelapse**	**To test the efficiency of the weeding algorithm ; the collect data on plant growth.**	**Chatelain & IAAC**	**Collected**	**Project**	**Collected and created data**	yes	**Researchers, farmers, developers**	**Tables and docs, pictures**	**PDF, JPG, MPEG**	**PDF reader, image viewer, video player**	**< 100 GB**	**Public, consortium**	**Image database system**	**Google Drive or project server**	**Metadata (yes) DOI (no)**	**No**	**10 yrs**
FARM-RESULTS	Outdoor Studies Results	Analysis of the data that was collected and evaluation of the		Chatelain, IAAC, SONY, CNRS	Created	Project	Analysis results		Researchers, farmers	Docs and tables	PDF	PDF reader	< 100 MB	Public	Google Drive	Google Drive	NA	No	10 yrs
Plant modelling																			
MODEL-CODE	Plant Models Code	Computational models of weeds and crop		INRIA & CNRS	Created	Project	Source code		Researchers, developers	code	L-Py, Python	Text editor	< 10 MB	Public	Inria disks	Cloud-based code repository (Github)	NA	No	10 yrs
MODEL-RENDER	**Plant Models Images**	**Synthetic images, rendered from virtual plants or 3D scans**	**Used to train learning machines for plant trait extraction**	**INRIA, CNRS & SONY**	**Created**	**Project**	**Collected and created data**	yes	**Data scientists, developers, researchers**	**Synthetic images**	**TIFF, JPEG**	**Image viewer**	**< 1 TB**	**Consortium, public**	**Inria disks and/or consortium's image database**	**Google Drive or project server**	**Metadata (yes) DOI (no)**	**No**	**Min. 5 years**
3d reconstruction of plants																			
SCAN-CODE	3D Image Processing Code	Algorithm to process 3D data for 3D plan reconstruction		SONY, INRIA, CNRS	Created	Project	Source code		Research groups working on plant	code	Python, C++	Text editor	< 50MB	Public	Local git repository	Cloud-based code repository (Github)	NA	No	10 yrs
SCAN-SYSTEM	3D Image Processing Pipeline	The image pipeline (data transfer, storage, retrieval) for the scanning		SONY, INRIA	Created	Project	source code & design files		Research groups working on plant	Source code & design documents	Javascript, Python, C/C++	Text editor	< 10 MB	Public	Local git repository	Github	NA	No	5 yr
SCAN-DATA	**Scanning Data**	**Data obtained using the plant scanner: Images captured for selected plants in indoor/outdoor settings using various 2D and 3D**	**To evaluate the algorithms of the scanner. To obtain data of plant growth.**	**SONY, INRIA, CNRS**	**Collected**	**Project**	**Collected and created data**	yes	**Research groups working on plant reconstruction**	**Images, point clouds, mesh, relative camera positions**	**Imaging data: JPEG, MPEG, PCD, PLY; meta-data: JSON, CSV**	**See list of software inaccompanying text**	**< 10 TB**	**Public**	**Local NAS or image database system**	**Google Drive or project server**	**Metadata (yes) DOI (no)**	**No**	**5 yrs**
Extraction of characteristic plant parameters using A.I. techniques																			
AI-CODE	Machine Learning Code	Code of the AI components to extract plant traits.		INRIA, Sony & CNRS	Created	Project	Source code		Application developers, researchers	code	Python, C/C++	Text editor	< 10 MB	Public	Local git repository	Cloud-based code repository (Github)	NA	No	10 yrs
AI-RESULTS	Machine Learning Results	The results and the evaluation of the A.I. algorithms.		INRIA, Sony & CNRS	Created	Project	Analysis results		Application developers, researchers	Docs, tables and images	PDF, JPEG, CSV, PLY, PCD, JSON	See list of software inaccompanying text	< 10 TB	Public	Local disks (Inria, SONY)	Google Drive	NA	No	10 yrs
Active vision to maximise the efficiency of the 3d scanning																			
ADAPSYS-DESIGN	Active Systems Designs	Descriptions of the harware sensors and experimental set-up used to run		UBER	Created	Project	Design files		Researchers in robotics, application	Docs	PDF	PDF viewer	< 100 MB	Public	Local disks	Cloud storage (Google Drive and/or	NA	No	10 yrs
ADAPSYS-CODE	Adaptive Systems Code	Code of the adaptive system algorithms		UBER	Created	Project	source code		Researchers in robotics, application	code	Python, C++	Text editor	< 10 MB	Public	Local git repository	Cloud-based code repository (Github)	NA	No	10 yrs
ADAPSYS-DATA	**Adaptive Systems Data**	**Data gathered during the experiments (development, testing) of the adaptive systems**	**To evaluate exploration algorithms.**	**UBER**	**Collected**	**Project**	**Collected and created data**	yes	**Researchers in robotics, application**	**Images, videos, meta-data**	**JPEG, MPEG, JSON**	**See list of software inaccompanying text**	**< 1 TB**	**Public**	**Local disks**	**Google Drive or project server**	**Metadata (yes)**	**No**	**5 yrs**
ADAPSYS-RESULTS	Local server (CNSR & Sony)	Results and evaluation of the adaptive system experiments		UBER	Created	Project	Analysis results		Researchers in robotics, application	Docs, images, videos, meta-data	PDF, JPEG, MPEG, JSON	See list of software inaccompanying text	< 100 MB	Public	Local disks	Google Drive	NA	No	10 yrs
In-field and in-door phenotyping to study plant development																			
PHENOTYPE-DESIGN	Phenotyping Design	Description of experimental set-up of the phenotyping studies including the description of plants developmental characteristics (Bolting time, flowering time, number of rosette leaves/branches, flowers/fruits,		CNRS	Created	Project	Design files		Any research labs interested in Arabidopsis development	Docs	PDF	PDF viewer	< 100 MB	Public	CNRS local server	Cloud storage (Google Drive and/or Github)	NA	No	10 yrs
PHENOTYPE-CODE	Phenotyping Code	Specific code used to analyse the data collected during the phenotyping experiments.		CNRS & INRIA	Created	Project	source code		Any research labs interested in Arabidopsis	code	R-code, Python	Text editor	< 10 MB	Public	CNRS local server	Cloud-based code repository (Github)	NA	No	10 yrs
PHENOTYPE-DATA	**Phenotyping Data**	**Data obtained through the manual scoring from laboratory experiments on Arabidopsis**	**Obtain detailed phylotaxic data of the Arabidopsis.**	**CNRS & INRIA**	**Collected**	**Project**	**Collected and created data**	yes	**Any research labs interested in Arabidopsis**	**Docs, measurements, images, video**	**PDF, CSV, JPEG, TIFF, MPEG**	**See list of software inaccompanying text**	**< 1 TB**	**Public**	**CNRS local server**	**Google Drive or project server**	**Metadata (yes)**	**No**	**To be defined**
PHENOTYPE-RESULTS	Phenotyping Results	Results and evaluation of the phenotyping experiments.		CNRS & INRIA	Created	Project	Analysis results		Any research labs interested in Arabidopsis	Docs with figures and images ; video	PDF, MPEG	See list of software inaccompanying text	< 100 MB	Public	CNRS local server	Google Drive	NA	No	10 yrs
Monitoring the plant growth to optimise the harvest management																			
GROWTH-CODE	Plant Growth Code	Specific code used to analyse the plant growth.		SONY, CNRS & INRIA	Created	Project	source code		Researchers in agronomy	code	Python, R-code	Basic office softwares, image and video viewers	< 10 MB	Public	Local server (CNSR & Sony)	Cloud-based code repository (Github)	NA	No	10 yrs
GROWTH-DATA	**Plant Growth Data**	**Plant growth data. This dataset may overlap with the other datasets (FARM-DATA, SCAN-DATA, PHENOTYPE-DATA).**	**The datasets used to analyse the plant growth and develop indicators on plant/crop status for**	**SONY, CNRS & INRIA**	**Collected**		**Collected and created data**	yes	**Researchers in agronomy, farmers**	**Measurements, images, video**	**JPEG, TIFF, MPEG, JSON, CSV**	**See list of software inaccompanying text**	**< 1 TB**	**Public**	**Local server (CNSR & Sony)**	**Google Drive or project server**	**Metadata (yes) DOI (no)**	**No**	**10 yrs**
GROWTH-RESULTS	Plant Growth Results	Results and evaluation of the plant growth analysis.		SONY, CNRS & INRIA	Created	Project	Analysis results		Researchers in agronomy, farmers	Docs with figures and images ; video	PDF, MPEG	See list of software inaccompanying text	< 100 MB	Public	Local server (CNSR & Sony)	Google Drive	NA	No	10 yrs
Exploitation																			
MARKET-INTERVIEWS	Market data	Data and information collected during the market studies, including recordings and transcriptions of interviews with farmers.	Obtain insights into the market of agro-robotics. Obtain insights on the best approaches to	FEI, SONY & IAAC	Collected	Project	Analysis results		Industry, government, EU	Recordings, texts, spreadsheets	PDF, MP3	See list of software inaccompanying text	< 1 TB	Consortium	Google Drive	Google Drive	NA	No	10 yrs
Management																			
PROGRESS DATA	Progress data	Data related to the progress of the project (number of deliverables provided, financial statement, spent	Dataset used for illustrating the project progress	FEI	collected	partners	Analysis results		no external dissemination	Excel	Excel	Excel	<1MB	Consortium	Google drive	Google drive	NA	Yes	during the project
Communication																			
COMMUNICATION DATA	Communication data	Photos, videos, and audio recordings that document the ROMI project and visual material for presentations	Dataset used for communication activities (booklet, website ...)	All partner	collected /created	Project / partners	Media files		Press	Photos, videos, audio recording	JPEG, TIFF, MPEG, MP3	See list of software inaccompanying text	<10 MB	public	Google drive, partner website	Google drive, project website	no	no	10 years