WeLASER

Eco-innovative weeding with laser

Sustainable Weed Management in Agriculture with Laser-Based Autonomous Tools

D6.6 –Data Management Plan

Co-funded by the Horizon 2020 programme
of the European Union

[This page intentionally left blank]

Acknowledgement

WeLASER is a project co-funded by the Horizon 2020 Research and Innovation Programme of the European Union under the call “Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bio-economy” and the topic “Integrated health approaches and alternatives to pesticide use”

(H2020-SFS-04-2019-2020)

Grant agreement N. 101000256

[This page intentionally left blank]

Deliverable number	D6.6
Deliverable full title	Data Management Plan
Work-package / Task	WP6
Leader beneficiary	University of Bologna - UNIBO
Deliverable type *	Report
Dissemination level**	Public
Status –Version	V1
Contractual delivery date	31/03/2022 (M18)
Actual delivery date	31/03/2022 (M18)
Star date/end date of the project	October 1, 2020 / September 30, 2023
Author(s)/contributor(s)	Giuliano Vitali (UNIBO) with contribution of all other WeLASER partners
Approved by	Pablo Gonzalez-de-Santos (Project coordinador)

* Report; Prototype; Demonstrator; Other.

**Public; Restricted to other programme participants (including the Commission Services); Restricted to a group specified by the consortium (including the Commission Services); Confidential, only for members of the consortium (including the Commission Services).

[This page intentionally left blank]

EXECUTIVE SUMMARY

In this deliverable, we describe the Data Management Plan (DMP) with reference to the different kinds of Research Data handled and produced by WeLASER.

WeLASER Consortium has adhered to the H2020 ORDP (Open Research Data Pilot) convention with the EC, which explicitly caters for the delivery of a DMP, together with a timely and rapid distribution of project results, making them widely available and openly accessible according to FAIR[1] principles.

This deliverable describes the data management plan, establishing the policy regulating collection, management, sharing, archiving, and preservation of research data in the WeLASER project. Data regulated by the plan are all Research Data (RD) - both confidential and public - that are either managed or produced by WeLASER researchers (i.e., data, charts, pictures, and documents) -, and the part of data to be disseminated through repository as Data Sets (DS).

SCHEDULED UPDATES

Issue	Expected by project month (M)
D6.5 - Initial DMP	M6
D6.6 - Intermediate DMP	M18
D6.7 - Final DMP	M36

[1] FAIR: findable, accessible, interoperable and reusable

[This page intentionally left blank]

TABLE OF CONTENTS

[This page intentionally left blank]

PARTNER ACRONYMS

CSIC	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (ES)
FUT	FUTONICS LASER GMBH (DE)
LZH	LASERZENTRUM HANNOVER e.V. (DE)
UCPH	KOBENHAVNS UNIVERSITET (DK)
AGC	AGREENCULTURE (FR)
COAG	COORDINADORA DE ORGANIZACIONES DE AGRICULTORES Y GANADEROS INICIATIVARURAL DEL ESTADO ESPANOL (ES)
UNIBO	ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA (IT)
IETU	INSTYTUT EKOLOGII TERENOW UPRZEMYSLOWIONYCH (PL)
UGENT	UNIVERSITEIT GENT (BE)
VDBP	VAN DEN BORNE PROJECTEN BV (NL)

LIST OF ACRONYMS AND ABBREVIATIONS

AI:	Artificial Intelligence
CA:	Consortium Agreement
CATI:	Computer-Assisted Telephone Interviewing
CBA:	Cost Benefit Analysis
CC:	Creative Commons
D:	Deliverable
DPO:	Data Protection Officer
DS:	Data Set
ERDA:	Electrical Research and Development Archive
FAIR:	Findable, accessible, interoperable, and reusable
FGI:	Focus group interviews
GA:	Grant Agreement
GDPR:	General Data Protection Regulation
ICP:	Informed Consent Process
IPR:	Intellectual Property Rights
LCA:	Life Cycle Analysis
NO-SQL:	Databases non constrained to SQL standard
OAI-PMH:	Open Archives Initiative Protocol for Metadata Harvesting
RD:	Research Data
SLCA:	Social-Life Cycle Analysis
SQL:	Structured Query Language
WP:	Work Package: WPn = work package number

[This page intentionally left blank]

1 DATA SUMMARY

A growing global population demands increasing food production, which requires increasing use of pesticides and fertilizers. About 130 million tons of herbicides per year are used in Europe alone that persist in the environment, destroy non-target plants and beneficial insects for the soil and produce health effects in animals and humans – cancer, birth defects and endocrine disruption. Moreover, existing herbicides become more and more ineffective due to the evolution and spread of herbicide-resistant weeds. Substitution of herbicides by mechanical automatic systems is under study, but mechanical solutions contribute to deteriorate the soil properties, harm beneficial soil organisms and provide poor results for in-row weeding.

WeLASER aims at developing a clean weed control system based on integration of autonomous field vehicles, vision-based weed identification and laser technologies. Aboard of the autonomous vehicle, AI-vision system discriminates crops from weeds and detects the position of the weed meristem to point and activate the laser on them using a laser scanner. The autonomous vehicle is driven by a smart controller embedded into an IoT ecosystem to monitor working area and environment, collect and elaborate validation data about test and efficiency of the tool.

WeLASER requires intensive collaboration for co-design and development of subsystems, system integration and testing. Most of the tasks require to fill out design reports, laboratory and field experiment setup aimed at tool tuning and validation leading to wide data collections.

Because of the industrial exploitation meaning of the project, most of material produced during design is confidential and related to industrial exploitation of results (by specific tasks, see Section 8 - Annex V).

1.1 Data for the Working of the Project

Research Data is generated by Tasks carried out within Work Packages 1-5, generating two different types of sources of RD, as described below:

1 RD from WP1 is related to the monitoring of the equipment development, observed failures or shortcomings to be used to find corrections, and develop assessment of procedures and usage protocols (communication, dissemination, exploitation, and risks). This RD, that includes survey generated by focus group interviews (FGI) integrated with planned discussion and interviews, literature study, and a quantitative survey (Computer-Assisted Telephone Interviewing, CATI), are used to refine the objective of the project and tune relevant design aspects, development of the invention, introduction to the market and impacts. Information collected also include data to be used to perform Economical and Social LCA analysis. Data for LCA are integrated with already existing data of Ecoinvent Ver 3 database[2] and literature data. LCA methodology is

[2] Ecoinvent database

based on the ISO 14040:2009 standard. The raw data will be kept confidential. In publication and dissemination only aggregated or anonymized data are used. As most of LCA data are based on commercial data-bases they cannot be disseminated in open Data Sets.

2 RD from WP2 - WP5 are related to the technology design, development and integration of the laser-based weeding system (WP2), weed-meristem perception system (WP3), autonomous vehicle (WP4) and the industrial integration and evaluation (WP5).

- RD form Tasks 2.1, 3.1, 4.1 and 5.1 consist of survey results and feed-backs from partners and stakeholders. These tasks relate to WP1 activities (interviews and focus groups).
- RD from Tasks 2.2-2.4, Tasks 3.2-3.5, Tasks 4.2-4.5, and Tasks 5.2-5.3 contain design features, data and calibration sheets, characteristics of case studies, and pictures. About the latter, a distinction must be made between pure image data and labelled image data. Labelled image data sets (partly originated or merged from those of previous or concurrent projects) cannot be disseminated in open Data Sets before completing the plan of exploitation of project results.

  As most of contents of RD cannot be disseminated before completing the strategy of exploitation of results, they are only synthesized in confidential deliverable, i.e., D2.1(by FUT), D3.1 (by LZH), D4.1 (by UNIBO), D5.1 (by CSIC), and D5.2 (by AGC). This research data will be stored in partner's data infrastructures for the whole duration of the project and the successive years specified in GA, under the supervision of researchers who are responsible of the data.

  Relevant research outcomes (e.g., features and usage of weeding implement based on plant experiments) are presented in scientific publications and, when not critical for future exploitation of the project, the underlying Data Sets are made openly available.
- RD from Task 2.5 and Task 5.4 consist of tables of (numerical and categorical) data, images and footage resulting from of dose-response experiments with weeds and crops, dose-response experiments with non-target small organisms, risk assessment of large organisms, and field test on selected crops to investigate the impact on crop productivity, soil quality and climate-change mitigation effects. Those RD are synthesized in the public deliverables D2.3 and D5.3, they are also presented in scientific publications, and deposited in repository as publicly available Data Sets. Such material of major importance to the objectives of the project is therefore useful at spreading WeLASER results for any future research.

A more detailed description of WPs and related Tasks is given in Section 8 - Annex V.

1.2 Research Data generation and processing

Research Data are represented both by qualitative data and quantitative data coming from several sources: meeting records & notes, questionnaires, experimental design, test & calibration trials, field robot trials, environmental and vehicle metering, together with pictures and footage.

RDs are collected by observation techniques related to specific lab & field experiments, in this case oriented to assess the efficacy of devices and weeder toolset (laser, pointer, scanner) developed by WeLASER (see Section 8 - Annex V).

As observation techniques undergo a continuous refinement, RD structure and contents are dynamically adjusted during the project.

RD are formerly collected / ingested / digitized, and successively validated, checked, cleaned-out and sorted to be selected for analysis (see below).

Data recorded from IoT framework are recorded in NO-SQL storage systems (data lakes) hosted by already available infrastructures at UNIBO.

The quality of the data is carefully assured using different approaches for the different data types - each single data (included open and check questionnaire, images and pictures) is time-space referenced.

Researchers and data responsible (see Table 4, Table 5), will detail methodology & protocol adopted, together with any relevant information on measurement devices and instruments characteristic (e.g., precision).

Research Data are regularly cleaned and validated to make them consistent. During this process any personal data - including images and footage with accidental inclusion of people - is 'wiped out 'before the deposit in repository and the dissemination.

Research Information and Data are communicated and shared by partners through cloud & storage systems with up-to-date privacy & security system management technologies chosen among public and commercial services.

Data containing personal information are not shared among partners, they are only temporary stored and backed-up on the local storage systems sited in EU of the researchers that are responsible for them, according to what indicated in the informed consent procedures and authorized through an informed consent form (see section "2.7 Ethical aspects").

2 FAIR DATA

This DMP follows the EU guidelines[3] and describes the data management procedures according to the FAIR principles[4] . The acronym FAIR identifies the main features that the project research data must have in order be findable, accessible, interoperable and re-useable, allowing thus for maximum knowledge circulation and return of investment.

In WeLASER a large amount of data is under production, made of environmental data at different

[3] Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016),

[4] The FAIR data principles (Force11 discussion forum)

spatial and time resolution, pictures, images and footage for monitoring crop grow, together with robot and tool activities. Such raw data are applied successive filtering processes depending on the objectives of technology development process, including (but not only) debugging, tuning, crop stage and mission based protocols. Only data considered functional to an analytic procedure are stored for further processing, under the responsibility of researchers involved and suitable to become openly available (see FAIR) and standardized Datasets (DSs) to be deployed in selected repositories (see table 1).

While Research Data could require a large amount of storage memory (data lakes), DSs should be bounded to a maximum size of 4Gb.

A list and description of Data Sets expected from WeLASER is given in section 3.

2.1 Making data findable, including provisions for metadata

At the moment of publication of project results, each research teams identifies the underlying data sets to be deposited together with the proper institutional or public long-time storage data repositories that can attribute persistent unique identifiers (DOI or Handle).

The chosen data repositories support standard descriptive metadata to ensure data sets indexing and discoverability - Dublin Core and Data Cite Metadata Schema. Moreover, they comply with the OpenAIRE 3.0 requirements for data archives. Consequently, the project data sets are visible via the OpenAIRE portal, facilitating project reporting procedures.

Table 1 reports the repositories for data sets publication and preservation chosen by WeLASER partners and the key features of the chosen repositories

Table 1. Summary of repositories.

Partner	Repository name	Type *(Institutional, Disciplinary, Multidisciplinary)*	Type of Permanent ID *(DOI / ULR / HANDLE, …)*	OpenAIRE level	Re3data
UGent IETU UCPH LZH	Zenodo	Multidisciplinary	DOI	OpenAIRE Basic (DRIVER OA)	https://www.re3data.org/repository/r3d100010468
UNIBO	AMS Acta	Institutional	DOI	OpenAIRE Data (funded, referenced datasets)	https://www.re3data.org/repository/r3d100012604
CSIC	Digital.CSIC	Institutional	HANDLE DOI	OpenAIRE 3.0 (OA, funding)	https://www.re3data.org/repository/r3d100011076

Partner	Repository name	Type *(Institutional, Disciplinary, Multidisciplinary)*	Type of Permanent ID *(DOI / ULR / HANDLE, …)*	OpenAIRE level	Re3data
UCPH	Electronic Research Data Archive (ERDA)	Institutional	DOI	not indexed in OpenAIRE	not indexed in Re3Data
LZH	Forschungsdaten-Repositorium der LUH	Institutional	DOI	not indexed in OpenAIRE	https://www.re3data.org/repository/r3d100012825

As some repositories are not (yet) indexed in OpenAIRE (UPCH, LZH one's), supplementary copies of DSs are to be deposited in Zenodo (see first row Table 1).

Specific keywords refer to thesauri and controlled vocabularies are associated to each data set to enhance semantic discoverability. Any specific ontology is mentioned in DS description file (README).

A nomenclature about DS is reported in Section 4 - Annex I

2.2 Making data openly accessible

WeLASER project does not have a data access committee, however as a guiding principle, WeLASER consortium seeks, whenever possible, to make RD openly available (as DSs), in order to achieve project's main aim, allow dissemination, validation and re-use of research results.

Restrictions to access are foreseen in the following cases:

- personal data of stakeholders involved in focus groups and meetings, interviews, and case studies,
- several WPs and Tasks are based on a collaboration with companies (both partners and third parties) that implies sharing know-how and industrial solutions that make research data confidential (see Section 8 - Annex V).

Nevertheless, all possible and legitimate actions and strategies are adopted to allow data sharing:

- during WeLASER focus groups and meeting with stakeholders only contact information are collected, and an informed consent are requested to stakeholders involved, while related data are properly anonymized (described in D8.2 - see Section 8 - Annex V for details),
- trying to obtain copyright permissions from third party data owners to be allowed to re-use, reproduce and distribute the collected data.

Research data that are structured and published as DS are converted from proprietary formats to well-known and documented, open, formats to facilitate accessibility and reusability (Table 2).

Versions or parts of the DSs that cannot be freely shared are indicated in Section 5 - Annex II, providing the specific motivations.

Table 2. Summary of data format.

Type of data	Formats used during data processing	Formats for sharing reuse and preservation
Textual data (e.g. Survey interviews)	TXT, XLSX	TXT, PDF, RTF
Bibliographic data	BIB, XLSX	BIB
Numerical data (e.g. tables & spreadsheets)	CSV, XLS, JSON	CSV, XLSX
Pictures	JPG, PNG	JPG, PNG
Videos	MPEG	MPEG

Datasets deposited in repositories are self-consistent and contain reusable data of interest to different potential users. For each deposited data set, all relevant documentation explaining data collection procedures and analysis (such as codebooks, methodologies, etc.) is made available along with the data to guarantee intelligibility, reproducibility, and the validation of the project findings. Moreover, the deposited documentation specifies the tools and software recommended to reproduce and reuse the data, when necessary. (See Table 3 for examples of tools and software enabling reuse of the data-set).

2.3 Making data interoperable

All data sets are described using standard descriptive metadata, such as Dublin Core and DataCite Metadata Schema, as required by the OpenAIRE Guidelines, to ensure metadata interoperability for indexing and discoverability. All relevant documentation explaining codebooks, users' manuals, data collection procedures and analysis are made available along with the data to guarantee intelligibility, reproducibility, and the validation of the project findings.

The chosen data repositories support protocols for the interoperability of metadata such as Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

To allow data exchange and re-use among researchers, institutions, organizations, countries, etc., partners convert all shareable data from proprietary formats and make them available in well-known and documented, open, formats (see Table 1 for details), as much as possible compliant with available (open) software applications. In case particular software is used in data processing, full explanation and instructions are included in the deposited documentation (a summary of the tools and software necessary to reuse of data sets is described in Table 3).

Table 3. Summary of tools and software for enabling re-use of the data sets.

Data type	Format	Tools/software required
Textual data	TXT, RTF, PDF	Open ASCII editors (e.g., Notepad++) Open world processors (e.g., OpenOffice, LibreOffice) Open PDF readers (e.g., PDF eXchange)
Bibliographic data	BIB	Open reference management software (e.g. Zotero)
Numerical data	CSV, XLSX	Open ASCII editors (e.g., Notepad++) Open spreadsheets (e.g., OpenOffice, LibreOffice)
Pictures	JPG, PNG	Open image editing and visualization tools (e.g. GIMP)
Videos	MPEG	Open video streaming tools
Compressed folders	ZIP	Open compression tools

2.4 Increase data re-use (through clarifying licences)

WeLASER distributes the shareable data by adopting licenses that allow re-use of the data in their entirety by scholars, stakeholders, and other users. The data sets are made available, unless otherwise stated, under Creative Commons (CC) BY 4.0[5] or CC0[6] upon each DS final contents[7].

In general, data are made openly available as underlying data necessary to validate the research results immediately at the time of publication of public reports and scientific papers. Data are given full citation from official project publications and web site, and they are made available through institutional or public data repositories compliant with OpenAIRE requirements[8]. (See Table 2).

The research data that cannot be shared because of confidentiality are those dealing with development of core technologies of the project, namely Laser, Laser pointing system, and field robot (dealing with most of deliverable of WP 2, 3, 4, 5). The strategies resulting from tasks T6.5 ("Management of Intellectual Property Rights") and T6.6 ("Plan for the exploitation of results") also detail how data are usable by third parties, after the end of the project. Confidentiality of produced DSs also reflects on deliverable, as described in Section 8 - Annex V.

[5] Creative Commons Attribution 4.0 International (CC BY 4.0),

[6]Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication,

[7]https://www.openaire.eu/how-do-i-license-my-research-data

[8] OpenAIRE, For Data Providers

2.5 Allocation of resources

Making data FAIR requires an investment of money and researchers' time. In WeLASER case, cost of data preservation after the project end are null because the chosen repositories do not apply fees for archiving and data curation.

During the project, a cloud service is going to be adopted to share IoT data among partners, under development at UNIBO. The cost to activate and maintain it for the duration of the project is covered by the project budget. The budget covers also the costs related to the project website setting up (from coordinator).

Costs related to data management and documentation, conversion of proprietary data files into open formats, and deposit procedures can be estimated about 3-5% of the number of Person-Months assigned to each Partner for the research activities.

A special case is represented by the time-consuming activities related to processing of interviews (i.e., transcription, translation, and anonymization). For each Partner involved, processing costs are estimated to be about 0.1 Person-Months/hour of processed audio recording, and they are already considered in budgets (COAG, IETU). Personnel dedication to WP6 goes from 1 PM (AGC, FUT), to 14 PM for UNIBO (including the activities related to the elaboration and update of the DMP).

2.6 Data security

The team leaders and their research collaborators (reported in Tables 4 and 5) are responsible for the management and storage of data .

Table 4. Summary and contacts of the research team leaders.

Team	RD responsible	ORCID ID (if available)
UNIBO	Vitali, Giuliano	0000-0002-7866-5534
CSIC	Gonzalez-de-Santos, Pablo	0000-0002-0219-3155
IETU	Krupanek, Janusz	
COAG	Mª García, Carmen	
Ugent	Xavier, Gellynck	0000-0002-8908-3310
UCPH	Andreasen, Christian	0000-0003-0844-141X
LZH	Wollweber, Merve	0000-0001-7024-2484
VDBP	van Zoggel, Paol	
FUT	Scholle, Karsten	
AGC	Aubé, Christophe	

Partners have identified all contributors (See Table 5) participating in data management activities, specifying their roles according to a given standard vocabulary (DataCite Metadata Schema).

Table 5. Summary of team members involved in the data sets collection and management.

Team	Member	ORCID ID (if available)	Role
UNIBO	Francia, Matteo	0000-0002-0805-1051	Researcher
	Arru, Marco		Researcher
CSIC	Emmi, Luis	0000-0003-4030-1038	Researcher
	Fernández, Roemi	0000-0003-0552-5407	Project member
IETU	Grzegorz Fronc		DataManager
	Katarzyna Lubera		DataManager
	Maryla Korcz-Olejek		DataManager
	Beata Michaliszyn- Gabryś		Project member
	Joanna Kulczycka		Project member
	Maria Buszman		Project member
	Marek Hryniewicz		Project member
	Mariusz Kalisz		Project member
UGent	Di Minh, Duc Tran	0000-0003-4249-9076	Project member
	Margo Degieter	0000-0002-4050-7485	Project member
AGC	Thomas de Saintignon		Project member
COAG	Roberto Saez		Project member
	Garau, Laura		Project member (to September.30.2021)
UCPH	Mahin Saberi	0000-0003-1456-1119	Project member
LZH	Hustedt, Michael	0000-0003-1991-8666	Project member
	Brodeßer, Alexander	0000-0003-1518-7501	Project member (to September.30.2021)
	Sandmann, Hendrik	0000-0001-9984-4745	Project member
	Hohndorf, Ruben		Project member (to September.30.2021)
	Ripken, Tammo	0000-0003-3512-9544	Project member (from October.1.2021)
	Worzischek, Malte		Project member (from October.1.2021)
	Lünsmann, Lorenz Alexander	0000-0002-4987-3860	Project member (to September.30.2021)
FUT	Anja Ahrens		Project member (to March.31.2022)
	Michael Schäfer		Researcher (from April.1.2022)
	Malte Kaule		Project member
VDBP	Jacob van der Borne		Project member

Keys for "Role" column: Data Collector (such as survey conductors, interviewers, etc.), Producer (person responsible for

the form of a media product), Project Member (a researcher indicated in the GA), Researcher (an assistant to one of the authors who helped with research, data collection, processing and analysis but is not part of team indicated in the GA), Research Group (the name of a research institution or group that contributed to the data set).

(See Section 5 - Annex II for details about data management responsibilities related to each project data set).

At each institution, research data are stored in computers, laptops, intranets or hard-drives (including RAID system) accessible through institutional password periodically modified according to national law provisions for data security and protected by regularly updated firewalls & antiviruses.

All the data are password protected. If mobile devices are used to store data files (e.g., backup files), they are kept in a safe place accessible only to the researchers involved.

All Partners are kept local updated copies of all their files: all the research materials stored in computers are subject to daily backup to safeguard them from accidental losses, that mostly also include regular backup on large data storage systems as SACO (CSIC cloud sistem for data-backup) and version management systems (GitLab) working in 'intranet' mode.

Long term preservation of public data is ensured by depositing the data sets (DSs) in repositories that have specific preservation policies. UNIBO AMS Acta, for example, guarantees long term preservation to the archived materials also thanks to a deposit agreement with the National Central Library in Florence. Zenodo policy ensures that the items are retained for the lifetime of the repository and in case of closure, best efforts are made to integrate all content into suitable alternative institutional and/or subject based repositories. ERDA guarantee the availability of the archived data and the UCPH-minted DOIs for at least 10 years.

2.7 Ethical aspects

WP8 is devoted to the ethical aspects related to the protection of personal data, the protection of the environment and the safety of the staff involved in the WeLASER project. Deliverables D8.1 and D8.2, which have already been submitted, have a direct impact on the DMP and must therefore be taken into consideration.

In deliverable D8.1, the procedures to be implemented for the collection, storage, protection, and destruction of personal data have been defined, and a Data Protection Officer (DPO) has been appointed for each of the tasks identified as requiring protection of personal data. All the procedures defined in D8.1 comply with the EU legislation. The template for informed consent and informed consent procedures have also been presented in D8.1. In addition, deliverable D8.2 explains the responsibilities of the DPO and includes their appointments in its Appendix A.

The technical, organizational and security measures to protect the rights and freedom of the data subjects, the techniques used for anonymization or pseudonymization and the explicit confirmation that the data used in the project are publicly available, as requested by the EC, are also addressed in D8.2. Besides the protection of personal data, we have not identified additional ethical issues.

In D8.1, it is stated that the link between the person's identity and the collected data are deleted one month after the collection (to guarantee anonymity). The questionnaires dealing with personal data include informed consent for data sharing, as well as information that the data collect could be used for further research beyond the end of the project.

3 DATA SET OVERVIEW

Table 6 offers an overview of the current state of the data sets expected from the project, to be adjusted and described in more detail in the final DMP. Each DS is detailed as reported in Annex I, II and III (Sections 4, 5 and 6, respectively).

Table 6. Data sets list.

n°	TASK	CT	DATA SET TITLE (contents)	SOURCE	STATUS
1	**5.2**	**UNIBO**	WeLASER. IoT in Crop Management. Survey	C	A
2	**5.2**	**CSIC**	Images of a wheat field in early growth stage	NYA	NYA
3	**5.2**	**CSIC**	Images of a maize field in early growth stage	NYA	NYA
4	**3.3**	**LZH**	TBD (image dataset without labels)	NYA	NYA
5	**1.1**	**COAG**	TBD* (Collection of data SH-interviews & meetings)	NYA	NYA
6	**1.3**	**UGENT**	TBD (Collection of data on Economic Assessment based on: data from partners, results of WP1, literature data)	NYA	NYA
7	**1.4**	**IETU**	TBD (Collection of data for E-LCA based on: data from partners, literature data)	NYA	NYA
8	**1.5**	**IETU**	TBD (Collection of data S-LCA based on: expert knowledge, partners, results from CATI & FGI, questionnaires, methodology literature general knowledge; generates: workshop reports, minutes, audio recordings of FGI & CATI)	NYA	NYA
9	**5.4**	**CSIC**	TBD (Platform Tests)	NYA	NYA
10	**5.4**	**UPCH**	TBD (Laser / Pointer lAB/ Field Test)	NYA	NYA
11	**5.4**	**UNIBO**	TBD (Environmental / Crop Monitoring)	NYA	NYA
12	**5.4**	**UNIBO**	TBD (Soil/Plants Tests)	NYA	NYA

* Table acronyms and abbreviations: TBD: to be defined, n°= data set progressive number, LB = WP lead beneficiary, PP = project phase (starting month-ending month), CT = creator team in charge of curating the data set, C=collected, G=generated, A=available, IP=in progress, NYA=not yet available.

4 ANNEX I: DATA SET NOMENCLATURE

WeLASER DSs are named following common rules to improve data visibility, discoverability, citation, and permanent online tracking.

The DS title structure is coded as:

WeLASER_Title specifying dataset content, coverage, and a version number

Example:

WeLASER_IoT-in-Crop-Management-Survey

The DMP also recommends the following rules for file naming:

– for data set file(s)

WeLASER_Keywords-specifying-coverage-and-nature-of-data. [ext] Example of spreadsheet .xlsx file:

WeLASER_IoT-in-crop-management-Survey.xlsx

– for README file (see Section 5 - Annex II)

WeLASER_ keywords-specifying-coverage-and-nature-of-data _README.ext Example:

WeLASER_IoT-in-crop-management-Survey_README.rtf

where [**ext**] is file extension. For README files suggested extensions are .txt (or any ascii-readable format, e.g., Markdown) and macro-free formats, e.g., Rich Text Format.

5 ANNEX II -"README" FILE TEMPLATE

A "README" file is a document that is deposited with each dataset, containing relevant information about data set authorship, terms of reuse and responsibilities, explaining data set content and structure, collection procedures and analysis (such as file specifics, methodologies, code-books of variables, data sources, and further necessary notes). The template of the README file that is used by WeLASER partners is shown here.

Introductory section

Data Set Title: "[insert title as defined in the DMP]"

Data Set Author/s: Name Surname (Affiliation), ORCID (if available);

Data Set Contributor/s: Name Surname (Affiliation), ORCID (if available);

Data Set Contact Person/s: Name Surname (Affiliation), ORCID (if available), email;

Data Set License: this data set is distributed under a (INSERT LICENSE)

Publication Year: (insert YEAR)

Project Info: [PROJECT ACRONYM] ([project full title], funded by European Union, Horizon 2020 Programme. Grant Agreement num. [grant agreement number]; [project website url]

Data set Contents

The data set consists of:

[Indicate the files that compose the dataset and their name and format.

In the following examples the data sets were composed by only one file. In case the dataset consists of more files you can name them as described and put them in a compressed folder. In this case readme file name should match the compressed folder name]

Data set Documentation

Abstract

[Insert a brief abstract describing the content of the dataset]

Content of the files:

file [Insert filename] contains ... [Provide a brief description of the content of the file/s. This is an example of how you could start]

File specifics

[Provide useful info regarding file conversion etc… (Optional)

Please indicate instruction/technical info to allow potential users to correctly visualize and reuse your data (e.g., specific software, …).

In case of data converted in open formats it could be useful to provide some further information. For example, if you deposit for long term preservation a .csv file derived from an excel you can describe the conversion. Here is an example of description of conversion using libre office calc software:

To create the .csv files, “LibreOffice Calc” version: 5.1.4.2 (portable) was used, with the following specifics:

- Character set Europa occidentale (Windows-1252/WinLatin1)
- Field delimiter « , » (comma)
- Text delimiter « “ » (quotes)]

Notes [Related to the whole dataset or to single files of a multi-file dataset (Optional)]

Data sources [Optional]

Methodologies [If necessary to understand how to reuse data]

6 ANNEX III: DATA SETS TABLE DESCRIPTOR

Analytic descriptions of template and currently deposited data sets of WeLASER project are reported hereafter.

Template:

DSN.n -not yet available - available -in progress	Dataset title
ID [ID type]	
Version	
Team in charge	
Creator/s	Family name, given name [TEAM]
Contributor/s	Family name, given name [TEAM]
Contact Person/s	Family name, given name [TEAM, MAIL]
Contents	
Data format	
Data volume	
Accessibility	Open, *restricted, closed* - license (e.g., CC BY) - (specify embargo)
Related publication/s	

#1 available	WeLASER. IoT in Crop Management. Survey
DOI	http://doi.org/10.6092/unibo/amsacta/6592
Version	v1
Team in charge	UNIBO
Creator/s	Vitali, Giuliano [UNIBO]; Francia, Matteo [UNIBO]; Golfarelli, Matteo [UNIBO]; Canavari, Maurizio [UNIBO]
Contact Person/s	Vitali, Giuliano [UNIBO, giuliano.vitali@unibo.it]
Contents	This dataset contains the underlying data of the following publication: "Vitali G, Francia M, Golfarelli M, Canavari M. Crop Management with the IoT: An Interdisciplinary Survey. Agronomy. 2021; 11(1):181. https://doi.org/10.3390/agronomy11010181". This data summarizes article citations aggregating them in categories and reporting the related keywords.
Data format	.xlsx, .bib
Data volume	49 kB
Accessibility	Data accessible under Creative Commons Attribution 4.0 International (CC BY 4.0) license.
Related publication/s	Vitali G, Francia M, Golfarelli M, Canavari M. Crop Management with the IoT: An Interdisciplinary Survey. *Agronomy*. 2021; 11(1):181. https://doi.org/10.3390/agronomy11010181

7 ANNEX IV: OPEN ACCESS STATUS OF PROJECT PUBLICATIONS

In the following table (Table 7) it is reported the updated list describing the open access status of the project publications related to data sets reported in Section 6 - Annex III.

Table 7. Open access status of WeLASER publications with indicated the related data sets.

Publications	
Vitali G, Francia M, Golfarelli M, Canavari M. Crop Management with the IoT: An Interdisciplinary Survey. *Agronomy*. 2021; 11(1):181. https://doi.org/10.3390/agronomy11010181	
Open Access repository URL	http://hdl.handle.net/11585/789000
Status	Open Access, indexed in OpenAIRE
Related dataset	Vitali, Giuliano ; Francia, Matteo ; Golfarelli, Matteo ; Canavari, Maurizio (2021) *WeLASER. IoT in Crop Management. Survey*. University of Bologna. http://doi.org/10.6092/unibo/amsacta/6592
Rakhmatulin I and Andreasen C. A Concept of a Compact and Inexpensive Device for Controlling Weeds with Laser Beams. *Agronomy*. 2020; 10(10): 1616.	
Open Access repository URL	https://www.mdpi.com/2073-4395/10/10/1616
Status	Open Access, indexed in OpenAIRE
Related dataset	to be deposited
Rakhmatuiln I, Kamilaris A, Andreasen C. Deep Neural Networks to Detect Weeds from Crops in Agricultural Environments in Real-Time: A Review. *Remote Sens*. 2021; 13(21): 4486.	
Open Access repository URL	https://www.mdpi.com/2072-4292/13/21/4486
Status	Open Access, indexed in OpenAIRE
Related dataset	to be deposited
Andreasen C, Scholle K, Saberi M. Laser Weeding with Small Autonomous Vehicles: Friends or Foes? *Front. Agron*. 2022;	
Open Access repository URL	https://www.frontiersin.org/articles/10.3389/fagro.2022.841086
Status	Open Access, indexed in OpenAIRE
Related dataset	to be deposited

8 ANNEX V: EXCERPT OF GA ABOUT WP'S TASKS

WP1 "Open-ended multi-actor networking and activities: from initial specifications to exploitation". This pool of activities will follow an interdisciplinary and multidimensional approach to deal with multiple effects in different domains, forecasting system behaviour and technology evolution, uncertainties, and risks. Related tasks are recalled hereafter.

TASK 1.1 (TL: COAG) "Identification, involvement, coordination and knowledge exchange with stakeholders and other entities" point to select stakeholders to be involved into focus groups and workshops to assess: i) technical, functional and economical aspects of the development and application of the new technology; ii) social and behavioral, legal and system conditions affecting farmers' adoption of the innovative technology; iii) environmental impact of the innovative technology and the requirements concerning labour, health safety and risk management in farms (see Task 1.5).

TASK 1.2 (TL: CSIC) "Scientific and technical continuous assessment – Value chain follow-up" is devoted to monitoring all the aspects included in the value chain, regarding both scientific-technical and evaluation issues, as much as monitoring marketing opportunities. These steps will also involve stakeholders and institutions for feedback.

TASK 1.3 (TL: UGENT) "Economic assessment and risk management in farms" focus on development of a model for the investment profitability assessment (on the level of individual farmer and entrepreneur) and development of cost benefit analysis (CBA).

TASK 1.4 (TL: IETU) "Health and environmental issues" regards health and environmental issues assessment of WeLASER technology through the Life Cycle Assessment (LCA) methodology based on the ISO 14040:2009 standard.

TASK 1.5 (TL: IETU) "Social aspects concerning the adoption of novel techniques" fill focus on Social Life Cycle Assessment (S-LCA) to assess the social and socio-economic aspects of innovative product and their potential positive and negative impacts along with its life cycle encompassing manufacturing; distribution; use; re-use; maintenance; recycling; and final disposal.

Research data collected during Task 1.1 (Subtask 1.1.1) will be used to refine the objective of the project and tune relevant design aspects, further development of the invention and its introduction to the market. RD from Subtask 1.2.1 will be used to monitor the equipment development to detect possible failures or shortcomings and suggest/collect corrections, while those from subtask 1.2.2 will be addressed to the assessment of procedures (communication, dissemination, exploitation, and risks). RD collected from partners during Task 1.3 (Subtask 1.3.1) will be used to perform an analysis including identification and evaluation of investment and operational costs (life cycle perspective) with evaluation of economic benefits to farmers, considering opportunities for innovative economic models like machine sharing/leasing/lending, and accounting for the geographic heterogeneity (biological, geological, structural, environmental) within the EU in the agriculture sector. RD from

T1.4 (S1.4.1) collected from WeLASER partner/partners and integrated with Ecoinvent Ver. 3 and literature data to compare WeLASER solution to existing agricultural practices in weeding (chemical, mechanical).

The results obtained will be described in deliverables D1.1 to D1.3 (public) and oriented to identify the potential of improvement of the social sustainability of the products. They will be used for dissemination and internal and available for any further development. Results from Task 1.4 will also contribute to deliverable D8.3 (EPQ-Requirement No. 3) and used for dissemination and for internal and available for any external further development. RD from Task 1.5 (Subtask 1.5.1) include qualitative survey generated by focus group interviews (FGI) integrated to planned discussion and interviews, and literature study, and a quantitative survey (CATI) as an initial stage of research leading to a wider qualitative analysis.

WP2 - WP5 are dealing with technology design, development, and integration:

WP2 "Laser-based weeding system: Development and impact" is aimed at design, development, tests, and evaluation activities.

WP3 "Weed-meristem perception system"

WP4 "Autonomous vehicle for laser weeding"

WP5 "Industrial integration and evaluation"

Involved activities are regularly subject to survey and feedback from partners and stakeholders (multi-actor involvement procedure), in **Task 2.1 - Task 5.1** (TL: IETU) reported in annual deliverables (D1.1 – D1.3) and with Research Data already discussed above (Task 1.1).

Other involved tasks are briefly recalled, and related RD discussed hereafter.

TASK 2.2 (TL: FUT) "High-power laser source" is oriented to development of a new fibre laser source to supply the laser scanner.

TASK 2.3 (TL: LZH) "Laser-based weeding scanner" is aimed at development of the device for directing and positioning the laser beam to precisely hit the weed's sensitive area (meristem).

TASK 2.4 (TL: FUT) "Weeding system integration and technical evaluation - TRL" is aimed at Individual tests for the system components (subtask 2.4.1), weeding system integration (subtask 2.4.2) check of the integrated weeding system (subtask 2.4.3).

RD collected during lab and field test of tasks 2.2-2.4 contains data and calibration sheets, features of experiments and case study. Such RD are confidential and will be synthesised in D2.1 (by FUT, confidential).

TASK 2.5 (TL: UPCH) "Impact of laser doses on living organisms" is devoted to experimentally study the effect of different laser doses on target and non-target living organisms and involves Dose-response experiments with weeds and crops (subtask 2.5.1), dose-response experiments with non-

target small organisms (subtask 2.5.2), risk assessment of large organisms (subtask 2.5.3) and field test on selected crops (subtask 2.5.4). This task will origin the majority pf public RD from WP2. These RD are of major importance to the objectives of the project. They will be synthesized in D2.3 (public), will be used to produce publication, and they will be useful for any future research aimed at enhancing WeLASER results, researchers, and developers. RD will consist of raw data and images data at the base of reports, statistics, tables, and other elaborated information data.

TASK 3.2 (TL: LZH) "Weed-meristem perception device" will be developed to identify target plants, target's meristem, and to provide a 3D localization of the meristem for the weeding scanner.

TASK 3.3 (TL: LZH) "Crop/Weed discrimination algorithms" focus on the design and development of the recognition software for precise detection and classification of crops and weeds in the intra-row space.

TASK 3.4 (TL: LZH) "Impact-point AI-vision system and weeding control system" is devoted to the design and development of the innovative AI-vision system to detect and precisely localize

the sensitive growth centres of plants (meristems).

TASK 3.5 (TL: CSIC) "Perception system integration and TRL evaluation" entailing the Individual tests for the perception device (subtask 3.5.1), weeding system integration (subtask 3.5.2), TRL assessment (subtask 3.5.3).

RD of tasks 3.2-3.5 contains data and calibration sheets, features of experiments and case study. Such RD are confidential and will be synthesised in D3.1 (by CSIC, confidential).

TASK 4.2 (TL: AGC) "Adaptation of the mobile platform" focus on the adaptation of the autonomous vehicle by AGC to support the laser-based weeding tool.

TASK 4.3 (TL: CSIC) "Smart central controller" devoted to developing the smart central controllers, and in particular the development of control unit (subtask 4.3.1), Input/output module (subtask 4.3.2) and system integration (subtask 4.3.3).

TASK 4.4 (TL: UNIBO) – "IoT and cloud computing –integration and management" aimed at developing and test field IoT devices (subtask 4.4.1), on board IoT devices (subtask 4.4.2) and cloud system (subtask 4.4.3).

TASK 4.5 (TL: CSIC) "Autonomous vehicle integration and TRL evaluation" aimed at Individual tests for the system components (subtask 4.5.1), autonomous vehicle integration (subtask 4.5.2) and TRL assessment (subtask 4.5.3).

RD of tasks 4.2-4.5 contains data and calibration sheets, features of experiments and case study. Such RD are confidential and will be synthesised in D4.1 (by UNIBO, confidential).

TASK 5.2 (TL: CSIC) "Equipment integration and testing" is oriented at a series of activities: system breakdown, integration procedures, test scenarios and assessment criteria (subtask 5.2.1), mechanical and communication integration check (subtask 5.2.2), final integration (subtask 5.2.3).

TASK 5.3 (TL: CSIC) "Equipment evaluation and TRL assessments" devoted to conduct the final tests and experiments, evaluating the performances of the final equipment and determining that the TRL is achieved.

RD of tasks 5.2-5.3 contains data and calibration sheets, features of experiments and case study. Such RD are confidential and will be synthesised in D5.1 (by CSIC, confidential), and D5.2 (by AGC, confidential).

TASK 5.4 (TL: UCPH) "Impact of the weeding equipment on crops and soil" has the scope of study of the impact on crop productivity (subtask 5.4.1), the impact on soil quality and climate-change mitigation (subtask 5.4.2). RD produced in this task are of major importance to the objectives of the project. They will be synthesized in D5.3 (public), will be used to produce publication, and they will be useful for any future research aimed at enhancing WeLASER results, researchers, and developers. RD will consist of raw data and images data at the base of reports, statistics, tables, and other elaborated information data.

WP6 - "Knowledge spread and innovation management" includes task about regulation of "multi-actor involvement procedure" (Task 6.1, TL: IETU), communication activities (Task 6.2, TL: COAG) leading to “products about the Launch of the project website and social media channels and accounts” (D6.1, by CSIC), and practice Abstracts (24 of them will be issued each year, documented in deliverable, and reported on web-site). Data will be publicly accessible via the project web site at https://welaser-project.eu, and related data reside on a server managed by administrative personnel of coordinator (at CSIC). The server is backed up weekly. Data will be maintained at least five years after the life of the project. Task 6.3 (TL: UPCH) is about dissemination of results leading to deliverable “Communication, dissemination and exploitation activities and results” (D6.2, D6.3, D6.4 by COAG). Task 6.4 will be about production of DMP (D6.5, D6.6, D6.7, by UNIBO), Task 6.5 on "Management of Intellectual Property Rights" (TL: CSIC), Task 6.6 (TL: UGENT) on "Plan for the exploitation of results" leading to D6.8 (by CSIC). Finally, T ask 6.7 (TL: VDBP) will focus on "Exhibitions, on-farm demonstrations and training activities" leading to D6.9 (by COAG).

WP7 is about coordination and will deliver regular reports on "risk management actions" each 6-months.

WP8 focus on "ethics requirements" that the project must comply with and will be discussed below along with description of FAIR policy.WP6, WP7, WP8 are not producing any RD.