D6.4 Data Management Plan

WP6 – Project Management
Version 1.0

Action Number: 101059784
Action Acronym: CROPINNO
Action title: Stepping Up Scientific Excellence and Innovation Capacity for Climate-Resilient Crop Improvement and Production
Work Programme: HORIZON-WIDERA-2021-ACCESS-02
Project Start: 09/01/2022
Project End: 08/31/2025

Deliverable Lead: IFVCNS
Dissemination Level: Public

Deliverable due date: 28/02/2023
Actual submission date: 28/02/2023
Doi: 10.5281/zenodo.7674225

HISTORY OF CHANGES			
Version	**Publication date**	**Main changes**	**Page (in the relevant version)**
1.0	28/02/2023	- Initial version	

Table of Contents

Executive Summary

This document made use of the Horizon Europe Data Management Plan Template (Version 1.0 dated 05 May 2021) and was written with reference to the Guidelines to FAIR data management in Horizon Europe. This Deliverable D6.5 is the initial Data Management Plan (DMP) for the CROPINNO project, funded by the EU's Horizon Europe Programme under Grant Agreement number 101059784. The purpose of the DMP is to provide an overview of all datasets collected and generated by the project and to define the CROPINNO consortium's data management policy that is used with regard to these datasets. It includes all necessary processes and procedures to ensure and cover the full life cycle of scientific data and research results generated during the project. It describes the different types and nature of the data including the procedures on how they are being collected, stored and shared. The DMP guarantees that the data follow the FAIR data principle, meaning that data is findable, accessible, interoperable and re-usable. Furthermore, the CROPINNO DMP lays out the procedure for data collection, consent procedure, storage, protection, retention and destruction of data, and confirmation that they comply with national and EU legislation. It reflects the status of the data that is collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. It follows the guidelines of the European Commission on FAIR Data Management in Horizon Europe as described in article 17 and analyzed in the Annotated Grant Agreement, article 17. This initial version of the DMP defines the general policy and approach to data management in CROPINNO that handles data management related issues. Being a living document, it will evolve, be revised and updated throughout the project's runtime, and the following versions will refine and enhance policy aspects and will go into more detail regarding the datasets collected and produced by the CROPINNO project.

Project Summary

The CROPINNO project establishes collaboration network between Institut za ratarstvo i povrtarstvo, Institut od nacionalnog značaja za Republiku Srbiju (IFVCNS), Novi Sad, Serbia and internationally-recognized research institutions from Spain (CSIC-IAS, Cordoba), Italy (UNIPD, Padova), and Germany (FZJ, Juelich and UROS, Rostock). IFVCNS is the National Institute of the Republic of Serbia and world-wide recognized institution in plant breeding. However, it still lacks in technology and innovation capacity in multi-omics and modern phenotyping, as well as in international visibility and project management and administration skills. The main objective of CROPINNO project is to step up and stimulate scientific excellence and innovation capacity of IFVCNS in the field of climate-smart crop improvement and production and enhance its ability to respond and create innovative solutions for the challenges that agriculture faces - climate changes and need to feed the increasing population. The other CROPINNO objectives are strengthening of the research management and administration skills of the IFVCNS and creating the conditions for positioning of IFVCNS as a regional hub of R&I in the area of agriculture and creation of Climate Crops Centre. These objectives will be achieved through a set of training, networking and dissemination activities, including short-term scientific missions, workshops, international summer schools, national and international conferences, as well as Field days. The expected impacts of CROPPINO include improved excellence capacity, enhanced strategic networking activities, raised reputation, research profile and attractiveness of IFVCNS and the research profile of its staff, strengthened research management capacities and administrative skills of IFVCNS staff, as well as improved creativity supported by the development of new research activities and collaborations and increased mobility of qualified scientists.

Keywords: Crop improvement; Climate change; Multi-omics; Phenotyping; Innovation capacity

Consortium Overview

This Data Management Plan was written by the CROPINNO project consortium which was formed with a goal of stepping up research capacity of IFVCNS in state-of-the-art techniques and protocols that would advance the existing crop breeding programs. IFVCNS National Institute of the Republic of Serbia, is a leading research center in classical crop breeding in the region, while it is severely lacking in knowledge and experience in application of modern tools for crop breeding and creating new generation climate-resilient crops. Universität Rostock (UROS) from Germany (Department of Plant Genetics) is well known for molecular sunflower hybrid breeding and well-developed system for drought tolerance research in potatoes which will be a model used for knowledge-transfer to other crops. Università di Padova (UNIPD) from Italy (Department of Agronomy, Animal Food Natural Resources and Environment) has research excellence in use of epigenetics and "–omics" data for crop improvement. Forschungszentrum Jülich (FZJ) from Germany (Institute for Bio- and Geosciences: IBG-2: Plant Sciences) is a world leading centre for the development and application of non-invasive sensors and automation technology to quantify structural and functional traits of shoots, roots, up to populations. Agencia Estatal Consejo Superior De Investigaciones Cientificas - Instituto de Agricultura Sostenable (CSIC) from Spain as well recognized excellence in integrated pest management for crop improvement. All four institutions will help IFVCNS tackle the major issues in crop breeding and designing resilient crops. All beneficiaries will participate in training (workshops, training schools, STSMs) of IFVCNS researchers and stakeholder engagement events. The knowledge and experience built up in the past decades by these leading institutions in numerous EU, Plant Phenotyping, and bilateral projects will be used to strengthen the outcome and impact of the CROPINNO project, putting the WIDENING Western Balkan institution IFVCNS at the forefront of knowledge and innovation in using modern phenotyping and multi-omics tools for creating more resilient crops for improvement of sustainability of agricultural production in Serbia and Western Balkans.

N°	Short	Beneficiary	Country
1	IFVCNS	Institut za Ratarstvo i Povrtarstvo, Institut od Nacionalnog Znacaja za Republiku Srbiju	Serbia
2	UROS	Universität Rostock	Germany
3	UNIPD	Università Degli Studi di Padova	Italy
4	FZJ	Forschungszentrum Jülich GmbH	Germany
5	CSIC	Agencia Estatal Consejo Superior de Investigaciones Cientificas - Instituto de Agricultura Sostenable	Spain

1. Data Summary

The table below shows the different collections of datasets that we intend to generate during the project. Furthermore, an overview of the naming conventions and versioning control of the datasets is given.

Datasets	Deliverable number	Dataset name	Work package number	Lead beneficiary or provider
Dataset 1	D1.4	Standardized protocols for efficient high-throughput phenotyping	WP1	FZJ
Dataset 2	D1.4	Standardized protocols for efficient genotyping	WP1	UNIPD
Dataset 3	D1.4	Standardized protocols in crop improvement programs	WP1	IFVCNS
Dataset 4	D2.2	Experiment: Phenotyping using a non-invasive phenotyping platform	WP2	FZJ
Dataset 5	D2.2	Experiment: Phenotyping using thermal imaging and other sensing techniques	WP2	CSIC
Dataset 6	D2.2	Experiment: Analysis of effects of drought and plant recovery at chromatin and transcriptional level	WP2	UNIPD
Dataset 7	D2.2	Experiment: Analysis of effect of drought using whole genome and transcriptome data for SNP analysis	WP2	UNIPD
Dataset 8	D2.3	Experiment: Integrated drought omics data	WP2	UROS

Consistent file naming of datasets is essential to guarantee their accessibility and interoperability. File names should be both machine and human readable. The following naming conventions are established between the beneficiaries:

- Start by identifying the project by the acronym in capitals (CROPINNO)
- Name your files consistently, keep the name short but descriptive (e.g. Report on agrifood sector needs)
- Use no special characters or spaces except underscores (e.g. Report_on_agrifood_sector_needs)
- Use consistent date formatting YYYYMMDD following ISO 8601 (e.g. for 20 May 2024 the format is 20240520)
- Insert version number of the dataset (e.g. v1.1, v2.1)

Therefore, the naming convention implies that the dataset file will be named as follows: CROPINNO_Report_on_agrifood_sector_needs_20240520_v1.1

1.1. Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

The existing data about root and shoot phenotypes and development of different crop species will be re-used to develop protocols for non-invasive phenotyping of sunflower genotypes under drought stress within the CROPINNO project. The re-used data will be used as reference data and will help to reduce the experiments for protocol developments to a minimum. Furthermore, linking the new data of sunflower with existing data of different crop species will give novel insights in general mechanisms in plant adaptations to stress environments. For data processing an existing open-source software (Zendro v3 open _ source software available under https://github.com/Zendro-dev/graphgl-server-model-codegen/blob/master/LICENSE, subject to the terms and conditions of GPLv3) will be implemented.
In addition, along the duration of the project, if there arises a need for re-using other existing data generated by other third parties, the beneficiaries will conduct all necessary checks to make sure that the type of reuse we have in mind is permitted and does not breach any copyright or other legal restrictions.

1.2. What types and formats of data will the project generate or re-use?

The data types in the datasets are qualitative and quantitative, textual and numerical, digital and digitized, raw and processed, and might include images and video data.
During data generation, collection and processing, the formats used are the ones which users are comfortable and used to working with, including proprietary formats. However, using non-proprietary data formats will reduce the risk of inaccessibility and obsolescence in the long term. Therefore, before sharing and depositing, the data will be converted to the recommended open formats which allow exchange of data in a standardized way, as follows:

- Textual qualitative data: Rich Text Format (.rtf) plain text, ASCII (.txt)
- Tabular quantitative data with minimum metadata: comma-separated values (.csv)
- Tabular data with complex metadata: SPSS portable format (.por)
- Digital image data: TIFF 6.0 uncompressed (.tif), JPEG2000
- Digital video data: MPEG-4 (.mp4)

1.3. What is the purpose of the data generation or re-use and its relation to the objectives of the project? What is the expected size of the data that you intend to generate or re-use?

WP2 is structured as an exploratory research project and is dedicated to implementation of multi-omics and phenotyping tools for increased climate resilience of sunflower, chosen model crop for the project that will be exposed to different stressors, such as drought and biotic stressors – *Orobanche cumana* and *Macrophomina phaseolina*. The techniques implemented and validated on sunflower within the research project will be transferred into improvement programs of other field crops bred at IFVCNS.
The expected size of the data to be generated at UROS is as follows: potato whole genome sequences 1.4 TB, RNASeq sunflower 120 GB, sunflower genome sequences (if produced in time) 700 GB. The data volume shared by FZJ is estimated to be around 100 GB, including raw data (sensor or camera outputs) and analyzed data. The data volume shared by CSIC is estimated to be around 50 GB, including raw data (sensor or camera outputs) and analyzed data.

1.4. What is the origin/provenance of the data, either generated or re-used?

Raw NGS sequencing data produced by UNIPD in WP2 will be processed and analyzed mainly by using publicly available software and pipelines by UNIPD and Sequentia Biotech. The information about the origin of the data generated by the automated phenotyping facilities at Forschungszentrum Jülich GmbH is provided in the metadata.

1.5. To whom might your data be useful ('data utility'), outside your project?

The NGS sequencing data might be useful to the scientific community working with plants, sunflower in particular. All information regarding data collection and processing will be published when raw data and processed data are made available to the scientific community by publication and deposition in public repository following the requested metadata standards. The phenotypic data are useful for plant breeding, for example the selection of genotypes or plants with improved abiotic stress tolerance. It is planned to deposit the sequencing data upon publication into the NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/). Primer sequences and SNP analyses will be published in peer-reviewed international journals in immediate and full open access.

2. FAIR data

2.1. Making data findable, including provisions for metadata

2.1.1. Will data be identified by a persistent identifier?

Datasets generated at FZJ will be stored in the Phenomis database at Forschungszentrum Juelich GmbH and published in Zenodo repository which assigns DOI as the persistent identifier. Datasets generated at IFVCNS will be deposited in the institutional repository FiVeR which assigns handle as the persistent identifier. DIGITAL.CSIC assigns Digital Object Identifiers (DOI) to datasets for their clear identification and citability and fulfils all the quality criteria established by Re3data directory of repositories. DOIs will be assigned to generated datasets directly by DIGITAL.CSIC through CSIC institutional membership to DataCite once the datasets are uploaded into its platform. In addition, DIGITAL.CSIC gives a handle identifier to any output that is uploaded onto its platform. Like DOI, handle is another global and permanent identifier for research outputs.
UROS will use e!DAL repository which assigns DOI.

2.1.2. Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will befollowed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Rich metadata will be provided with the data to allow discovery. The metadata include authorship, platform identifier and time of creation, defines the data formats, traits, treatments used, contains the context that led to the data, include settings of experimental sensors or instruments, environmental conditions, comments, and measurement uncertainties, etc. Metadata standard to be used is Dublin Core (https://www.dublincore.org/), README files will be provided alongside the data, keywords will be provided in the metadata to optimize discovery and re-use. In FiVeR, metadata will be made available via the OAI-PMH protocol for harvesting aggregators and via page-level meta tags for web browsers, search engines and reference managers. Recommended citations in the APA, Vancouver and Chicago citation styles will be available on the data set's landing page in the repository, and metadata export in RIS and BibTeX formats will also be supported.
In order to make them easily findable, the data and the corresponding metadata uploaded in DIGITAL.CSIC will be described following the basic principles of Qualified Dublin Core which is the by default metadata standard in this DSpace-based repository. Furthermore, DIGITAL.CSIC offers at item level the option to export the metadata records in DataCite format, as well as other international formats, such as RDF, MARC, MODS, csv, and bibtex. In order to promote standard citation, DIGITAL.CSIC has produced a description template for datasets that follows recommendations set by FORCE11 guidelines

(https://www.force11.org/datacitationprinciples). In addition, recommended description in this template follows recommendations set by DataCite Metadata Working Group. This template is available at http://digital.csic.es/handle/10261/81323. This metadata template has been further extended so as to include DataCite specific metadata properties, such as types of contributors. The metadata schema describing the data files will follow as standard the Dublin Core Metadata Initiative (DCMI). DCMI standard satisfies international standards (ISO standard 15836:2009, ISO CEI 11179) and has an official status from the World Wide Web consortium (W3C) and ISO 23950 norm. The metadata set consists of fifteen elements as follow; title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights.

2.1.3. Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use? Will metadata be offered in such a way that it can be harvested and indexed?
Yes.

2.2. Making data accessible

Repository:

2.2.1. Will the data be deposited in a trusted repository?
Datasets generated at FZJ will be stored in the Phenomis database at Forschungszentrum Juelich GmbH which provides the services to allow acquisition, access, security and long-term sustainability of service and funding. The data are assigned with unique identifiers. In addition, the data will be included in the trusted general-purpose open repository Zenodo.
Datasets generated at IFVCNS will be deposited into the trusted institutional repository FiVeR (https://fiver.ifvcns.rs). The repository is OpenAIRE validated (information visible on website), it is registered in OpenDOAR, and is harvested by CORE and BASE. The repository uses a DSpace-based software platform developed and maintained by the Belgrade University Computer Centre (RCUB). The software platform is compliant with the OpenAIRE Guidelines for Literature Repositories v3. Explicit information about the repository policies and services is available online at http://fiver.ifvcns.rs/Files/policy-fiver-en.html. The repository assigns handles and ensures that contents are accompanied by enough machine-actionable and standardized metadata (Dublin Core). Currently, the repository is in the process of registration at https://www.re3data.org/. Although not yet certified, the repository follows best practice in terms of long-term preservation, as explained in the repository policy.
The CSIC datasets will be stored in DIGITAL.CSIC (https://digital.csic.es), the open access repository of the Spanish National Research Council. All datasets in DIGITAL.CSIC are accompanied with their descriptive metadata and are described according to international standards and good practices in line with journal's data sharing policy requirements and open data mandates. DIGITAL.CSIC accepts open access, embargoed and closed access datasets however there is a strong preference for the first two options unless legitimate reasons prevent from sharing the results openly. In any case, metadata of all items in DIGITAL.CSIC are publicly available and ready for consumption by human users and third-party research infrastructures, search engines and aggregators. OpenAIRE harvests DIGITAL.CSIC contents on a regular basis and so do other Open Science aggregators such as DataCite Search, SHARE of Center of Open Science (COS), and Google Dataset Search. In addition, DIGITAL.CSIC is findable through common search engines like Google, Bing and Yahoo. Further, metadata schema used in DIGITAL.CSIC allows to fully document details about data files access requirements and the repository's content policy permits the upload of research software associated with the datasets.
UROS will make use of the public PGP repository e!DAL - Electronic Data Archive Library (https://edal.ipk-gatersleben.de/). e!DAL support the FAIR principles and requirements from funding agencies and journals. It supports metadata standards, like DublinCore or Schema.org. e!DAL is designed to support JSON-LD, OAI-PMH and ORCID allowing data harvesting and easy linkage across global services. e!DAL tracks the version history of all stored data objects to make the whole research process transparent and reproducible. It orchestrates pre-existing local file storage and embedded metadata database towards a comprehensive, big data aware storage backend. Its main features are version tracking, metadata management, information retrieval, journal and founding agency proven registration of persistent identifiers (DOI), an embedded HTTP(S) server for public data access, access as a network file system, and a scalable storage backend.

2.2.2. Have you explored appropriate arrangements with the identified repository where your data will be deposited?
Yes, all the identified repositories are free for use, either institutional or Zenodo.

2.2.3. Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?
Yes.

Data:

2.2.4. Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.
Research data resulting from the project will be deposited in the trusted repositories and made openly available as soon as

possible after the consortium has analyzed them and published their findings, following the guidelines set in the IP Management Plan D5.3., IP Exploitation Plan D5.4 and Communication and Dissemination Plan D5.1. Deposition of data will take place as soon as possible after data production/generation or after adequate processing and quality control have taken place, providing value and context to the data and at the latest by the end of the project, so that metadata information is available and hence information about the data is findable. Data underpinning a scientific publication will be deposited at the latest at the time of publication, and in line with standard community practices.
However, there will be different access levels. For instance, research results that can reasonably be expected to commercialization or industrial exploitation, and/or that require a special protection due to confidentiality will not be put into the open domain. Whenever such access restrictions apply the associated metadata available in DIGITAL.CSIC will contain the email contact of one member of the project consortium to ask for individual agreements on accessing the data, even if there are general restrictions. This action is compatible with the Grants agreement obligations (Article 29.3) concerning the dissemination of results, and in compliance with the obligation to protect results (Article 27), confidentiality (Article 36) security obligations (Article 37) and obligations to protect personal data (Article 39). Other datasets will be openly shared after an embargo period during which the project consortium will be able to examine the generated data and work on preliminary results. Metadata of such datasets will be publicly available on DIGITAL.CSIC unless they may compromise the outcomes of the ongoing investigation.
The research data will mostly be made available for re-use as soon as possible. Data used in a publication will be deposited on DIGITAL.CSIC as soon as the publication is made public by the journal or upon acceptance for publication, whatever the journal data sharing policy will state. The research data to be generated by the project and not directly associated with any publication will also be made available through DIGITAL.CSIC by the end of funded action.
In cases where data remain closed after careful consideration and decision of the project management, and there are no compelling reasons that the related metadata should not be findable and accessible, open access will be provided to the metadata of the data, with CC0 public domain dedication or equivalent, if possible, while the dataset itself remains closed.

2.2.5. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
Embargo on dataset access will be applied until articles are published or intellectual property protected to ensure exploitation of the results by the beneficiaries.

2.2.6. Will the data be accessible through a free and standardized access protocol?
The data will be accessible during and after the end of the project by converting the files to standard open formats. Data will be made available in a repository for dissemination, reports and publications, and they will be accessible through the standardized access protocol.

2.2.7. If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? How will the identity of the person accessing the data be ascertained?
DIGITAL.CSIC offers the option to embargo datasets for as long as needed. During the embargo period restricted access for private use may be requested at no cost through the “Request a copy” functionality available at the repository. The project will make use of such option for the data not directly associated with publications.

2.2.8. Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?
In case there are any issues regarding the restricted access to research results, CROPINNO`s Management Committee can act as data access committee and seek clarification.

Metadata:
2.2.9. Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?
Metadata of deposited datasets will be openly available under a Creative Commons Public Domain Dedication CC0 licence or equivalent to ensure its reusability, in line with the FAIR principles (in particular machine actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.
The data will be complemented by metadata that describe the research data. This includes authorship and time of creation, defines the data formats used and contains the context that led to the data. Additionally, metadata include settings of experimental sensors or instruments, environmental conditions, comments, and measurement uncertainties, etc. The metadata is essential for the re-use of research data ensuring also the possible verification of research results by third parties.

2.2.10. How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?
Data and metadata will remain available and findable indefinitely.

2.2.11. Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?
No documentation about a specific software is needed to access or read the data and metadata, as standard PC software is sufficient.

2.3. Making data interoperable

2.3.1. What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?
In line with the OpenAIRE Guidelines, the repository uses controlled values for the resource type, access level, licence and version fields. No discipline-specific vocabularies are currently supported.
FZJ will support IFVCNS in using data standards such as the public MIAPPE (https://www.miappe.org/) and BrAPI (https://brapi.org/) standards. Data hosted in separate warehouses will be integrated into a larger meta-dataset, a data-cloud. FZJ will also provide training on creating data-warehouses and how to reach the data management standards set in EPPN2020/EMPHASIS already existing tool sets that automatically create fully functional plug and play data warehouses, capable of distributed storage and data integration (https://zendro-dev.github.io/).
DIGITAL.CSIC by default metadata schema is internationally adopted Qualified Dublin Core and onsite conversion services for its metadata records are offered as far as DataCite schema, a global standard to describe research data and software, is concerned. The research datasets as well as associated metadata and documentation will be compliant to international standards in order to guarantee easy interoperability and re-use. The consortium will apply widely-used controlled vocabularies as needed.

2.3.2. In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?
Currently, CROPINNO does not intend to introduce new project-specific ontologies or vocabularies.

2.3.4. Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?
Yes.

2.4. Increase data re-use

2.4.1. How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?
Data documentation to be provided with datasets will include the README files, describing among the other things how all the files included within the dataset relate to one another. More specifically, the README files accompanying FZJ datasets will contain the information about the phenotyping platform and methods to measure and analyze the date, will define outlier detection, definition of traits and sensors used and units of measurements.
To ensure reproducibility and possible interoperability it is also crucial that every dataset is accompanied by a detailed documentation in form of the readme.txt file. This documentation will be submitted to DIGITAL.CSIC as a supplementary file together with the data files and will provide the necessary insights for future researchers about methodologies, approaches, decision making and software used.
Last but not the least, it is worth mentioning that DIGITAL.CSIC is about to complete the SCHOLIX standard and DataCite Event Data implementations, which will allow automatic linkage via an URL between the project's generated datasets to be available on DIGITAL.CSIC and associated publications. Such interoperability will enrich maximize discoverability, understanding, citation and impact of datasets.

2.4.2. Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?
Yes, all data that the will be open will also be feely available and released under the CC BY license in line with the obligations set out in the Grant Agreement.

2.4.3. Will the data produced in the project be useable by third parties, in particular after the end of the project? Will the provenance of the data be thoroughly documented using the appropriate standards?
Generated datasets, especially the ones with molecular and phenotypic data, will be disseminated and available to the wider scientific community, which will create opportunities to develop new collaborations and boost existing partnerships. Data provenance will be appropriately documented.

2.4.4. Describe all relevant data quality assurance processes.

Data collected manually are double-checked by team members based on random samples. Digitalized data are regularly checked for errors, duplicates or inconsistencies by creating a filter in .xls file. The filter allows researchers to see all of the unique values in the column, making it easier to isolate the incorrect values. Another quality assurance process includes selecting the cells or column you want to validate, selecting Data Validation on the Data tab, and selecting the kind of data that should be in the column in the Allow box. Options include whole numbers, decimals, lists of items, dates, and other values. For phenotyping reference targets and R scripts are used to check the accuracy / quality of the automated generated data.

2.4.5. Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

The CROPINNO project will generate other research outputs, namely the standardized protocols which will be treated regarding the allocation of resources, data security and ethical aspects in the same manner as other data stemming from the project and in accordance with the FAIR principles.

3. Other research outputs

3.1. In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Standardized protocols for stress application and digital workflow for studying plant response to drought stress and recovery from the stress, at transcriptional and chromatin level will be available to the scientific community following publication of results in open access journals. Unpublished information will be shared among partners, as stated in CA.

DOI minting through DIGITAL.CSIC is also offered to CSIC researchers for other non-traditional research outputs, such as research software, preprints and lab notebooks.

3.2. Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

The CROPINNO project will manage and share other research outputs, or make them available for re-use, in the same manner as other data stemming from the project and in accordance with the FAIR principles.

4. Allocation of resources

4.1. What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?

There are no immediate costs anticipated to make the open results generated in the project FAIR. Especially no costs are foreseen for storing open results in the projects chosen repositories Zenodo and other institutional repositories where institutional support and provisions are available. Additional details will be reported, as needed, in the following versions of this DMP.

UNIPD NGS raw data (RNA-seq reads and ChIP-seq-reads) will be stored and retained in in-house repository servers without any additional charge. Raw NGS sequencing data will be made fully available in public repository when articles are accepted for publication in immediate and full open access journals.

Regarding costs, DIGITAL.CSIC repository provides its services for free to institutional researchers and is maintained by the Unit of Scientific Information for Research of the Spanish National Research Council. Likewise, DIGITAL.CSIC mints DOIs for datasets, software and other non-traditional outputs free of charge for CSIC researchers.

4.2. How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

The costs for research data/output management, such as metadata generation, format conversions and providing description and documentation, as well as preparation of this DMP, will be covered by the project budget. At UNIPD, choice on the number of genotypes, biological and technical replicates for efficient NGS sequencing data integration of genomic transcriptomic and epigenomic analyses depends on the budget and data processing limitation.

Publication fees in full and immediate open access venues for peer-reviewed scientific publications will also be covered by the project funding.

4.3. Who will be responsible for data management in your project?

Project coordinator (Dragana Miladinović (IFVCNS), ORCID: 0000-0001-9555-9162) and vice-lead (Ankica Kondić Špika (IFVCNS), ORCID: 0000-0003-3298-6620) are responsible for:

- Developing the data management plan and policies in cooperation with WP leads
- Monitoring data management activities and deadlines
- Providing support to WP leads regarding data management

Work Package leads (Renate Horn (UROS), ORCID: 0000-0003-1174-8658, Serene Varotto (UNIPD), ORCID: 0000-0001-5219-7157, Kerstin Nagel (FZJ), ORCID: 0000-0003-3025-0388, Leire Molinero Ruiz (CSIC), ORCID: 0000-0001-5152-0418) are responsible for:

- The implementation of the data management policies in their respective WPs
- Monitoring data management activities and deadlines
- Ensuring that publications, datasets and metadata are properly managed and deposited in line with this DMP

Research staff, laboratory or technical staff (Laura Junker-Frohn (FZJ), ORCID: 0000-0002-0655-6232; Carmen Gómez-Lama Cabanás (CSIC), 0000-0001-8312-9253; María José Giménez-Alvear (CSIC), 0000-0003-3854-5746; Montserrat Jurado-Expósito (CSIC), 0000-0003-2796-0565) are responsible for:

- Collecting, processing and analyzing the data
- Generating, collecting, processing and describing the data in accordance with this DMP
- Generating metadata and documentation to accompany data

4.4. How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

UNIPD NGS raw data (RNA-seq reads and ChIP-seq-reads) will be stored and retained in in-house repository servers without any additional charge. The costs for long-term preservation of FZJ data will be covered by institutional funding within the Helmholtz Association (Germany). The datasets generated at IFVCNS will be preserved indefinitely in the institutional repository without any costs to the project.

Long term preservation by the DIGITAL.CSIC repository will result in no additional costs for the project consortium. The research data will be stored locally at the Spanish National Research Council with redundant, robust and secure storage. No additional costs are associated with this local storage. In addition, DIGITAL.CSIC disposes of long term storage, easy access and downloading facilities for open access data. The repository is an application under the umbrella of Central IT Services in the Spanish National Research Council and is included in the institutional preservation strategy. As a consequence, regular backups of DIGITAL.CSIC take place on a daily basis on institutional servers and copies of files are being kept in several offsite locations managed by CSIC. For its part, the Unit responsible for the repository is putting in place its own preservation system for all its library-related applications and the repository software DSpace has a number of tools to conduct regular audits in order to check the validity of formats and bitstreams and block any possible virus or undesired bot. It is worth mentioning that DIGITAL.CSIC obtained a Data Seal of Approval in 2015.

The CROPINNO Management Committee decides what data will be kept and for how long. The data have potential value for further plant breeding efforts.

5. Data security

5.1. What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Following the best practices, data is kept secured by making three back-ups in two locations (e.g. researcher`s computer, USB drive or external hard drive, and remote in-house approved space).

UNIPD NGS raw data (RNA-seq reads and ChIP-seq-reads) will be stored in in-house repository servers at DAFNAE and in backup storage spaces with automatic backup, before and after data processing. UNIPD will be responsible for management of NGS raw data (RNA-seq reads and ChIP-seq-reads) after sequencing and during data processing that will be done in collaboration with Sequentia Biotech.

The data stored at Phenomics database of Forschungszentrum Jülich will be backed up on a regular daily basis.

Data at UROS will be saved on an external hard drive in addition to the internal hard drive of the work station and on the server Synology NAS at the University of Rostock. At UROS the CLC Genomic Workbench will be used for the data analyses. RNAseq data will be deposited upon publication in the NCBI Sequence Read Archive (SRA) for public use. Metadata on the phenotyping will be made available via the public PGP repository e!DAL.

CSIC internal network will make backups of the files every day. It remains possible to retrieve all stored versions for the last weeks. CSIC retention policy secures files for years.

Open results deposited in the Zenodo repository are stored in CERNs EOS service (http://eos.web.cern.ch/content/about-eos). Each file copy has two replicas located on different disk servers. For each file, two independent MD5 checksums are stored. The servers are managed according to the CERN Security Baseline for Servers. For more information see http://about.zenodo.org/infrastructure/.

5.2. Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes.

6. Ethics

6.1. Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the

Description of the Action (DoA).

The research within CROPINNO project does not involve any human participants nor gathering or use of any personal data. Therefore, there are no considerations regarding participant identity or subsequent anonymization of personal data.
However, as outlined in the chapter 4 of the Description of Action (DoA) titled Ethics Self-Assessment, the project plans for the exchange of plant materials, in particular sunflower DNA, plant tissue, and germplasm among beneficiaries. The project will use IFVCNS proprietary sunflower inbred lines that will be exchanged with the partners in WP1 and WP2. Regarding compliance with relevant legislations and legal issues, material exchange will be organized formally by Material Transfer Agreement (MTA), and import rules, including phytosanitary regulations, will be followed where applicable.
No ethical issue is present. There are no personal data included and materials are not confidential.

6.2. Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

There will be no questionnaires dealing with personal data and no informed consent for data sharing within the CROPINNO project.

7. Other issues

7.1. Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

IFVCNS makes use of its institutional Rulebook on Open Science from 2021 (in Serbian Cyrillic, https://ifvcns.rs/wp-content/uploads/2021/07/Pravilnik-o-otvorenoj-nauci.pdf) aimed at implementing the Open Science Platform of the Ministry of Education, Science and Technological Development of the Republic of Serbia, primarily through the digital institutional repository FiVeR (http://fiver.ifvcns.rs/). In addition to general provisions on open access, Creative Commons licenses and submitting research outputs to the repository, the Rulebook gives detailed guidelines on submitting primary research data, FAIR principles, and discusses different options and types of repositories for data deposition.