

Grant agreement number: 101134965 Horizon Europe - Research and Innovation Action

Deliverable 5.2

Data Management Plan

Due date of the deliverable: M6 - February 2025 Actual submission date: M6 - February 2025

WP concerned	WP5			
WP leader	Jacques LE GOUIS	INRAE		
Author(s)	Michael ALAUX	INRAE		
Version V1	V1			
Dissemination level	PU: Public - fully open			



Funded by the European Union (Grant agreement nr 101134965), the Swiss State Secretariat for Education, Research, and Innovation (SERI, contract nr 24.00389) and UK Research and Innovation (UKRI, reference nr 10121521). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU, SERI or UKRI. Neither the EU nor the granting authorities can be held responsible for them.



Funded by the European Union

Table of Contents

1.	Sur	nmary		1
2.	Res	sults		1
2	.1	Plan Details	1	
2	.2	Project Details	1	
2	.3	Legal and ethical requirements, codes of conduct	4	
2	.4	Germplasm passport information	5	
2	.5	New phenotyping information	8	
2	.6	New genotyping information	11	
2	.7	Whole genome sequence	15	
3.	Add	litional communication & dissemination data considerations		.17
4.	Cor	nclusion		. 18



1. Summary

This deliverable consists of the first version of the PRO-WILD Data Management Plan (DMP). It has been created using DMP OPIDoR, based on the "Science Europe: structured template" provided by DMP OPIDoR.

An additional section (Chapter 3) covers communication and dissemination-specific data requirements, and was created separately.

This data management plan has been reviewed and validated by the members of the PRO-WILD executive committee (WP leaders, deputy leaders and species leaders).

2. Results

Data Management Plan of the project "Protect and promote Crop Wild Relatives"

A Data Management Plan created using DMP OPIDoR, based on the "Science Europe: structured template" provided by DMP OPIDoR.

Plan title	DMP of project "Protect and promote Crop Wild Relatives"
Deliverable	D5.2
Version	First version
Plan purpose/scope	
Fields of science and technology (from OECD classification)	1.6 Biological sciences, 1.7 Other natural sciences, 1.2 Computer and information sciences
Language	eng
Creation date	2025-02-04
Last modification date	2025-02-19

2.1 Plan Details

2.2 Project Details

Project	Protect and promote Crop Wild Relatives
title	



Acronym	PRO-WILD			
Abstract	Crop wild relatives (CWR) are wild plant species genetically related to cultivated crops. Their untapped diversity can enhance resilience to biotic and abiotic stress and the nutritional quality of modern crops. Wheat, sugar beet, and oilseed rape were selected by PRO-WILD because of their importance to food security and EU farmers and because some of their wild relatives are endemic to Europe. Moreover, these CWRs constitute a rich and under-exploited resource needed to face challenges linked to climate change and the transition to low-input agriculture. The genetic diversity and vulnerability of these CWRs must be better characterized to optimize their conservation and utilization. The objectives of PRO-WILD are to identify priorities for in-situ conservation of the selected CWR gene pools, to survey and complement CWR genebanks collections, and to increase the use of CWRs in crop improvement. PRO-WILD associates 18 partners from 11 EU and associated countries with expertise in ecology, conservation, genomics, pathology, microbiology, plant breeding, agriculture, and sociology. PRO-WILD will compile and analyse CWR occurrence maps and perform new in-situ collections. It will predict the vulnerability of several CWR species and populations to ongoing climate changes. Ex-situ collections will be complemented with endangered CWRs accessions. PRO-WILD will investigate the resilience of CWR collections for relevant biotic and abiotic stresses. The identification of CWR-derived desired traits and their transfer into elite backgrounds will be coordinated with input from breeders, farmers, and consumers. PRO-WILD outcomes will contribute to European Green Deal initiatives through research, education, and training. It will serve the EU biodiversity and the Farm to Fork strategies by preserving, characterising, and utilising wild species that have unique			
Funding	• European Commission : https://doi.org/10.3030/101134965			
Start date	2024-09-01			
End date	2029-08-31			
Partners	 SESVANDERHAVE KWS SAAT SE & CO KGAA ELLINIKOS GEORGIKOS ORGANISMOS - DIMITRA LEIBNIZ - INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG WEIZMANN INSTITUTE OF SCIENCE 			



THE AGRICULTURAL RESEARCH ORGANISATION OF ISRAEL - THE VOLCANI CENTRE
INRAE TRANSFERT SAS
UNIVERSITA DEGLI STUDI DI PALERMO
• INNOLEA
JOHN INNES CENTRE
FORSCHUNGSINSTITUT FUR BIOLOGISCHEN LANDBAU STIFTUNG
INSTITUT NATIONAL D'ENSEIGNEMENT SUPERIEUR POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT
UNIVERSITAET FUER BODENKULTUR WIEN
HIGHCLERE CONSULTING SRL
UNIVERSITY OF CUKUROVA
UNIVERSITA DEGLI STUDI DI PERUGIA
CONSIGLIO PER LA RICERCA IN AGRICOLTURA E L'ANALISI DELL'ECONOMIA AGRARIA
VEREIN DER ZUCKERINDUSTRIE EV

Research outputs:

1.	Germplasm passport information	(Dataset)
----	--------------------------------	-----------

- 2. New phenotyping information (Dataset)
- 3. New genotyping information (Dataset)
- 4. Whole genome sequence (Dataset)

Contributors

Name	Affiliation	Roles
Alaux Michael	INRAE	Contact Person (Passport data, Genome sequence data, Phenotypic data, Genotyping data) DMP manager
Le Gouis Jacques	INRAE	Project coordinator



2.3 Legal and ethical requirements, codes of conduct

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Description	PRO-WILD is committed to Open Science, and it will implement not only mandatory but also optional Open Science practices to make knowledge, methods, data and research outputs openly accessible to all, to strengthen scientific collaboration and knowledge sharing for the benefit of science and society and to open the knowledge production process to society, beyond the scientific community. INRAE has a policy and a roadmap for Open Science and, as the coordinator of PRO-WILD, it will ensure that good practices are followed.
	PRO-WILD will provide immediate Open Access (OA) to peer-reviewed scientific publications, including articles, book chapters, monographs, books and PhD theses. Authors will also be encouraged to provide open access to nonpeer-reviewed publications, which goes beyond the mandatory practice. Results will be published in open-access venues, such as journals (e.g. PLOS) and platforms (e.g. Open Research Europe). In addition, the author's final manuscript will be deposited in a trusted repository, institutional (e.g. HAL-INRAE, IRIS-UNIPG/UNIPA), domain-specific (e.g. Europe PMC) or general-purpose (e.g. Zenodo). The repository will inform on the research outputs, results, data, tools and instruments (e.g. software) needed to validate the conclusion of the publication. Where possible, open access to these research outputs and tools will be provided. Research outputs will be licensed under the most appropriate Creative Commons (or equivalent) open license according to their form. Beneficiaries and authors will retain sufficient intellectual property rights to comply with the open access requirements.
	In addition to mandatory open science practices, PRO-WILD will also implement the following optional practices: Early and open sharing of research. We will share results early by posting preprints of articles on bioRxiv and/or the project's open access repository (e.g. Zenodo), in line with the Dissemination Plan, provided this practice is accepted by the policy of the target journal.
	Participating in open peer review. When appropriate, PRO-WILD will consider seeking the recommendation of their preprints in Peer Community In (PCI Ecology, PCI Genomics), providing open access to the peer-reviews, author's responses and editorial decisions. PRO-WILD publications in Open Research Europe will also benefit from an open peer review.



2.4 Germplasm passport information

2.4.1 Data description and collection or re-use of existing data

Research output description

Name	Germplasm passport information
Description	PRO-WILD will use the list of Multi-Crop Passport Descriptors (MCPD), which is a reference tool developed jointly by the International Plant Genetic Resources Institute (IPGRI) and FAO to provide international standards to facilitate germplasm passport information exchange across crops.
Туре	Dataset
May contain personal data?	No

Will existing data be reused?

Justificatio	PRO-WILD initial survey of existing CWRs in partners' genebanks:					
n	number of accessions, methodology used to genotype/sequence (and associated projects).					
	Partner	Crop	Number of	Available	Available genome	
			accessions	genotyped	sequences and	
				accessions	(project)	
				methodology		
				(project)		
		Wheat	216	98 / 420k SNP	14 / Exome	
		Wheat	210	array	capture	
				(BreedWheat)	•	
				104 / 15K SND		
				array (AgroDiv)		
				-) (3)		
		Brassica	29		26 / Illumina 50X	
					3 / Illumina 100X	



WIS	Wheat	900	900 WEW / GBS	5 Aegilops species
				n partial Am. muticum
IPK	Wheat	1894	1894 / GBS (GeneBank3.0)	
IPK	Beet	900		264 / Illumina 7- 10X
				1 / assembled genome
				Sequencing performed by BOKU
ARO	Wheat	608		
ARO	Beet	115		
JIC	Wheat	609		242 whole- genome shotgun short-read
				sequencing (OpenWildWheat)
CREA	Wheat	450	260 WEW/ 35K SNP array	150 - 8X (OpenWildWheat)
CUK	Wheat	500		
ELGO	Wheat	211		
ELGO	Beet	70		
L				

2.4.2 Documentation and data quality

What metadata and documentation (for example way of organising data) will accompany the data?

Description	Persistent ID: DOI



	Data format : CSV	
	Standard : MCPD	
	Licence : CC-BY	
Metadata/data standards	 MCPD : Multi-Crop Passport Descriptor, https://www.fao.org/plant- treaty/tools/toolbox-for-sustainable-use/details/en/c/1367915/ 	

2.4.3 Data processing and analysis

How and with what resources will the data be processed / analyzed?

i.		
	Description	Each partner will collect and process the passport data of each accession.
		INRAE-URGI will provide the infrastructure to store and share the data between partners.
	Before the end of the project, each partner will contact their NFC (National Focal Point) to submit their passport data to Eurisco. INRAE-URGI could provide recommendations and support.	

2.4.4 Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	The data will be stored at the dedicated infrastructure hosted at INRAE-URGI to fill, curate and validate data. The data will be backed up as describe in the INRAE-URGI platform DMP: <u>https://doi.org/10.15454/9HM5UI</u>
Unit	MB
Equipment, technical platforms	 PlantBioInfoPF : https://cat.opidor.fr/index.php/PlantBioInfoPF



2.4.5 Data sharing and long-term preservation

How will data be shared?

Modalities of sharing	During the project, the validated data will be shared between partners using the INRAE-URGI infrastructure.		
	Before the end of the project, the validated data will be published in open repositories such as Zenodo and EURISCO.		
Reusability			
Data repository/catalogs	 Zenodo : https://cat.opidor.fr/index.php/Zenodo () EURISCO : http://eurisco.ecpgr.org () 		

2.5 New phenotyping information

2.5.1 Data description and collection or re-use of existing data

Research output description

Name	New phenotyping information			
DescriptionNew phenotyping information on important agroup physiological traits observed on different sets in controlled and in different field environments for the three CWR get About 50 experiments will be conducted for a total of ~15 and ~10,000 plants with on average 10 traits generating p and environmental data in CSV format (10-20 MB). Will be these experiments on two high-throughput phenotyping plat PlantArray (CREA) and 4PMI (INRAE) platforms (estim volume 1TB). 4PMI is part of the French research infr Phenome-Emphasis with its own storage capacity.Major PRO-WILD traits to phenotype on the different sp experimental networks involved.				
	Biotic interactions:			
	• wheats			
	Resistance to Zymoseptoria tritici (septoria leaf blotch), Fusarium Head Blight, leaf and yellow rust (WP3.3) - Germany, UK Root microbiome composition and exudate (WP1.4) - Israel, Turkey			



	Brassica		
	Resistance to Psylliodes chrysocephala (Cabbage stem flea beetle and Plasmodiophora brassicae (clubroot) (WP2.4, WP3.3) - Franc Germany		
	• Beets		
	Resistance to Beet Curly Top Virus (BCTV), and Syndrome des Basses Richesses (SBR) (WP2.4) - Germany		
	Abiotic constraints:		
	Wheats		
	Terminal drought tolerance in controlled conditions (Stomatal density and conductance, root morphology traits) (WP3.2) - Italy Field trials in different pedoclimatic zones under rainfed and irrigated conditions (WP3.2) - Greece, Italy, Turkey		
	Brassica		
	Functional traits (specific leaf area, chlorophyll content, etc.) related to adaptation to abiotic stress (WP1.2) - Italy		
	Beets		
	Drought and cold tolerance in field trials and under controlled conditions (WP2.4) - Greece, Italy		
	Nutritional value:		
	Wheats		
	Concentration and composition of proteins, mineral (zinc, iron		
	calcium) content, fibre and starch contents (WP3.2) - France, UK		
Туре	Dataset		
May contain personal data?	No		

How new data will be collected or produced?

Name of the method	



Description	Field trial experiments	
	Greenhouse experiments	
Data Nature	Experimental data	

2.5.2 Documentation and data quality

What metadata and documentation (for example way of organising data) will accompany the data?

Description	Persistent ID: DOI		
	Data format : CSV		
	Standard : MIAPPE, Crop Ontology (CO_321, CO_333)		
	Licence : CC-BY		
Metadata/data standards	 MIAPPE : Minimum Information About Plant Phenotyping Experiments, https://www.miappe.org/ 		

2.5.3 Data processing and analysis

How and with what resources will the data be processed / analyzed?

Description	n Each partner will perform their phenotyping experiments and store raw data.	
	INRAE-URGI will set up a dedicated MIAPPE compliant template.	
	Each partner will fill the template and cure their data until validation.	
	INRAE-URGI will provide the infrastructure to store the data and the automated tools that will help with the curation and validation of the data/metadata content.	
	INRAE-URGI will provide training and support to the tools and the template.	
	INRAE-URGI will help the validated data submission to the open repositories such as Eurisco and Zenodo before the end of the project.	



2.5.4 Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	The data will be stored at the dedicated infrastructure hosted at INRAE-URGI to fill, curate and validate data.			
	The data will be backed up as described in the INRAE-URGI platform DMP: https://doi.org/10.15454/9HM5UI			
Unit	GB			
Equipment, technical platforms	 PlantBioInfoPF : https://cat.opidor.fr/index.php/PlantBioInfoPF 			

2.5.5 Data sharing and long-term preservation

How will data be shared?

Modalities of sharing	During the project, the validated data will be shared between partners using the INRAE-URGI infrastructure.			
	At the end of the project, the validated data will be published in open repositories such as Zenodo and EURISCO.			
Reusability				
Data repository/catalogs	 Zenodo : https://cat.opidor.fr/index.php/Zenodo () EURISCO : http://eurisco.ecpgr.org () 			

2.6 New genotyping information

2.6.1 Data description and collection or re-use of existing data

Research output description



D5.2 - Data Management Plan

Name	New genotyping information
Description	New genotyping information by the most appropriate and cost- efficient methods (GBS) on ~4000 accessions that will produce approximately a volume of 30-50 GB. Genetic diversity analysis will yield classical VCF format files delivering the identified polymorphism and associated information (context, position, quality). The European Variation Archive (EVA), wheredata can be stored, is part of the ELIXIR infrastructure and is an open access database of all types of genetic variation data from all species.
Туре	Dataset
May contain personal data?	No

Will existing data be reused?

Justificatio n	PRO-WILD initial survey of existing CWRs in partners' genebanks: number of accessions, methodology used to genotype/sequence (and associated projects).				
	Partne r	Crop	Number of accession s	Available genotyped accessions and methodology (project)	Available genome sequences and methodology (project)
	INRAE	Wheat	216	98 / 420k SNP array (BreedWheat) 104 / 15K SNP array (AgroDiv)	14 / Exome capture
	INRAE	Brassic a	29		26 / Illumina 50X 3 / Illumina 100X
	WIS	Wheat	900	900 WEW / GBS	5 Aegilops species



					1 partial Am. muticum
	IPK	Wheat	1894	1894 / GBS (GeneBank3.0)	
	IPK	Beet	900		264 / Illumina 7- 10X
					1 / assembled genome
					Sequencing performed by BOKU
	ARO	Wheat	608		
	ARO	Beet	115		
	JIC	Wheat	609		242 whole- genome shotgun short-read
					sequencing (OpenWildWheat)
	CREA	Wheat	450	260 WEW/ 35K SNP array	150 - 8X (OpenWildWheat)
	CUK	Wheat	500		
	ELGO	Wheat	211		
	ELGO	Beet	70		



2.6.2 Documentation and data quality

What metadata and documentation (for example way of organising data) will accompany the data?

Description	Persistent ID: BioSample ID, EVA ID		
	Data format : VCF		
	Standard recommendations: FAIR cookbook		
	Licence : CC-BY		

2.6.3 Data processing and analysis

How and with what resources will the data be processed / analyzed?

Description	Each partner will perform their genotyping using GBS or other methodologies.					
	INRAE-URGI will set up the infrastructure to host the data to allow the sharing between partners.					
	At the end of the project, each partner will submit their data to EVA. INRAE-URGI could provide recommendations and support.					

2.6.4 Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	The data will be stored at the dedicated infrastructure hosted at INRAE- URGI to fill, curate and validate data.
	The data will be backed up as described in the INRAE-URGI platform DMP: <u>https://doi.org/10.15454/9HM5UI</u>
Unit	MB

2.6.5 Data sharing and long-term preservation

How will data be shared?



D5.2 - Data Management Plan

Modalities of sharing	During the project, the validated data will be shared between partners using the INRAE-URGI infrastructure.
	Before the end of the project, the validated data will be published in the European Variation Archive (EVA).
Reusability	
Data repository/catalogs	 EVA : https://www.ebi.ac.uk/eva/ ()

2.7 Whole genome sequence

2.7.1 Data description and collection or re-use of existing data

Research output description

Name	Whole genome sequence
Description	Whole genome sequences are to be obtained for ~50 wheat, ~10 Brassica and ~700 beet CWR for an estimated volume of 5-10TB. The European Nucleotide Archive (ENA) on which data can be stored is part of the ELIXIR infrastructure and represents an open, supported platform for the management, sharing, integration, archiving and dissemination of sequence data.
May contain personal data?	No

2.7.2 Documentation and data quality

What metadata and documentation (for example way of organising data) will accompany the data?

Description	Persistent ID: ENA Project-ID			
	Data format : BAM			
	Standard : Gene Ontology			
	Licence : CC-BY			
Metadata/data standards	 BAM : Binary Alignment Map, https://en.wikipedia.org/wiki/Binary_Alignment_Map 			



2.7.3 Data processing and analysis

How and with what re	esources will the	data be processed	d / analvzed?
now and with what it			a r unuiyzoù.

Description	Each partner will perform the sequencing and assembling their genomes.
	The assembled genome sequences will be stored on the INRAE-URGI infrastructure.
	A dedicated solution to share the data between partners will be studied (e.g. Globus, One Data).
	Before the end of the project, each partner will submit their genome sequences to ENA. INRAE-URGI could provide recommendations and support.
Equipment, technical platforms	 PlantBioInfoPF : https://cat.opidor.fr/index.php/PlantBioInfoPF

2.7.4 Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	The data will be stored at the dedicated infrastructure hosted at INRAE URGI to fill, curate and validate data.	
	The data will be backed up as described in the INRAE-URGI platform DMP: https://doi.org/10.15454/9HM5UI	
Unit	ТВ	

2.7.5 Data sharing and long-term preservation

How will data be shared?

Modalities of sharing	During the project, the validated data will be shared between partners using the INRAE-URGI infrastructure.
	Before the end of the project, the validated data will be published in the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home)
Reusability	



3. Additional communication & dissemination data considerations

•

This section ensures that PRO-WILD's communication outputs are properly managed, stored, and made widely accessible in line with Horizon Europe's open science requirements.

Data Description

PRO-WILD will generate and disseminate various types of materials to communicate project outcomes effectively. These include:

- Videos (e.g., project documentaries, stakeholder interviews, webinars) in MP4 format.
- Presentations (e.g., conference slides, training materials) in PPTX and PDF formats.
- Reports and policy briefs in PDF format.
- Infographics and visual materials in JPEG, PNG, and SVG formats.
- Social media content (text, images, and short videos) shared on platforms like Twitter, LinkedIn, and YouTube.

Data Storage & Backup

- All primary dissemination materials will be stored on PRO-WILD's institutional servers.
- Videos and larger files will be archived in Zenodo, the PRO-WILD website, and YouTube, ensuring open access.

Accessibility

- Reports, infographics, and presentations will be publicly available on the PRO-WILD website and Zenodo, with a DOI assigned for citation.
- Videos will be published on YouTube with subtitles for accessibility.
- Metadata (e.g., descriptions, keywords) will be added to all uploads to enhance discoverability.

Long-Term Preservation & Sustainability

• Key dissemination materials will remain available for at least five years post-project via institutional repositories and Zenodo.



- The PRO-WILD website will serve as a central hub, maintained by the coordinating institution for a minimum of three years post-project.
- If the website is decommissioned, all essential materials will be transferred to an open-access repository.

Ethical & Legal Considerations

- All materials adhere to GDPR: Any data featuring individuals (e.g., interviews, images) is collected and shared only with explicit consent.
- Third-party content (e.g., stock footage, external datasets) will only be used with proper licenses or permission.
- Attribution and citation guidelines will be followed for any referenced materials.

4. Conclusion

This DMP is a living document that will be updated throughout the life of the project. For example, ongoing discussions with the COUSIN sister project could lead to few updates.

