

TITLE of AGREEMENT	CWR in EURISCO
AGREEMENT NUMBER	L21ROM198
IMPLEMENTING PARTNER	Universidad Rey Juan Carlos
AUTHORS OF THE REPORT	José M. Iriondo, Ada Molina, Elena Torres, María Luisa Rubio, Guillermo Asens, Carlos Lara y Dolores Cuadra
DATE SUBMITTED	12/07/2023
TYPE OF REPORT	Interim Technical Report
ABSTRACT (Maximum 200 words)	<p>This interim technical report covers the activities carried out in the period January 1st – June 30th, 2023. The main achievements of this period include the completion of the deliverables ‘Identify populations from the CWR priority list of the National Strategy for CWR conservation’, which was due this semester (31/03/2023), and ‘Prepare a draft of the national database structure’, which is due on 30/09/2023. Moreover, significant advances have been made in the deliverable ‘Collect and organize available data according to the agreed principles and data exchange format’, specifically in terms of feeding content to specific descriptors of the database, such as those related with the administrative units in which the populations are located, elevation, biogeographic region, ecogeographic unit and whether the populations are located in protected areas. Finally, a tentative selection of 50 most appropriate sites for the establishment of genetic reserves have been identified, although further approaches to this selection are still pending.</p>
KEYWORDS	<p>Country/Region: Spain Crop(s): Crop Wild Relatives Subject: Extension of EURISCO for Crop Wild Relatives (CWR) <i>in situ</i> data and preparation of pilot countries’ data sets: Spain</p>

**Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and
preparation of pilot countries' data sets**

Spain

Interim Technical Report

Reporting period: January 1st – June 30th, 2023

Table of contents

Introduction	4
Deliverables due within the reporting period.....	4
Composition of the working team	4
Identify populations from the CWR priority list of the National Strategy for CWR conservation	5
Identify key public and private institutions for the organization of a national network of data providers	8
Prepare a draft of the national database structure	9
Collect and organize available data according to the agreed principles and data exchange format.....	13
Provide the data to EURISCO	14
References.....	14
Annex 1. National Inventory of Crop Wild Relatives and Wild Food Plants of Spain (CWR-V8). Logical database model documentation.....	16

Introduction

This interim report covers the activities carried out from 1st January to 30th June 2023 in the context of the project “[Extension of EURISCO for Crop Wild Relatives \(CWR\) *in situ* data and preparation of pilot countries’ data sets](#)” in relation to its implementation in Spain. The overall project aims to extend the [EURISCO catalogue](#) and prepare it to host and make the CWR *in situ* data provided by the European countries publicly available. It also intends to establish an agreed mechanism for data flow from national inventories to EURISCO, either through the existing National Inventory Focal Points or through a newly created network of specific CWR Focal Points. In parallel, a group of pilot countries is supported in the organization of their internal data flow and data gathering mechanisms, and also in the preparation of their *in situ* CWR data according to the standard requirements defined in this project. This new *in situ* leg of the EURISCO catalogue will be built in compliance with the ‘[FAIR principles](#)’ (findable, accessible, interoperable, and reusable data).

The activities carried out in the context of this project follow the recommendations indicated in the document “[Principles for the Inclusion of CWR Data in EURISCO](#)” agreed by project partners and the EURISCO Advisory Committee.

The key deliverables and activities to be implemented in the context of this project are:

1. Identify populations from the CWR priority list of the National Strategy for CWR conservation (31/03/2023).
2. Identify key public and private institutions for the organization of a national network of data providers (31/12/2022).
3. Prepare a draft of the national database structure (30/09/2023).
4. Collect and organize available data according to the agreed principles and data exchange format (15/11/2023).
5. Provide the data to EURISCO (15/11/2023).

Deliverables due within the reporting period

1. Identify populations from the CWR priority list of the National Strategy for CWR conservation (31/03/2023).

Composition of the working team

The AdAptA-Lab of the Rey Juan Carlos University has formed a team of researchers with expertise in conservation of crop wild relatives to participate in this project. The team is currently composed by José M. Iriando, Carlos Lara-Romero, Elena Torres, María Luisa Rubio Teso, Ada Molina, Dolores Cuadra, Guillermo Asens and Celia del Tío. Fernando Pomedá also worked on this project until the end of February 2023. He was contracted using the funds available from this project. His contract started on 26 September 2022.

Identify populations from the CWR priority list of the National Strategy for CWR conservation

The taxa selected to identify the populations were obtained from the Crop Wild Relative (CWR) / Wild Food Plant (WFP) priority list included in Spain's [National Strategy for the Conservation and Use of Crop Wild Relatives and Wild Food Plants](#), which was approved by the Ministry of Agriculture in July 2022 and published in November 2022. The Spanish CWR/WFP priority list contains 521 taxa delimited at the species level. The taxonomical nomenclature follows the ["Standard List" \(Lista Patrón\)](#) used by the Spanish public administration as a standard to be used in all public initiatives. This list is mostly based on the nomenclature used in *Flora iberica* (Castroviejo, 1986-2012).

The information to identify the distribution of the populations of the 521 selected taxa was obtained from [GBIF database](#) with an automated procedure developed under the R statistical environment (R Core Team, 2022), using 'rGBIF' package (Chamberlain & Boettiger, 2017). GBIF is the largest global repository of wild species occurrences and includes records from many national and international sources. Most Spanish institutions upload data on plant distribution to GBIF so this is the most efficient way of obtaining the plant distribution data available for Spain. 90% of the selected taxa matched the accepted scientific name of GBIF backbone taxonomy. In this case, the query was performed using the scientific name and GBIF provided the records belonging to these scientific names or any other synonyms. The remaining 10% of taxa were considered synonyms according to the GBIF taxonomic backbone. Because in several cases these taxa were made equivalent to a broader taxonomic concept (*i.e.*, they were considered subspecies of the accepted scientific name by GBIF), we queried the database using the original scientific names of the Spanish standard list and discarded the records that contained synonym scientific names according to GBIF. These queries to GBIF database for the 521 taxa provided 2,289,454 records.

To obtain a high-quality dataset to be used in subsequent analyses, the downloaded data were filtered using additional scripts developed by the working team in the R environment. In these scripts, several criteria were applied to remove low-quality records, including those without geographic coordinates, with low accuracy of the geographic coordinates, with errors in the geographic coordinates, associated with cultivation and/or botanical gardens, with incompatible land uses (water bodies, perpetual ice, urban settings), and with error-prone taxonomical determination. At the same time, all records with the same geographic coordinates or within a 500m radius buffer were given the same population identifier (POPID). The decision to incorporate all records that fall within a 500m radius has already been used in previous scientific research when there is no specific knowledge about the average pollination and seed dispersal distance corresponding to each of the target species (Iriondo, 2011).

During the semester corresponding to this interim report, the dataset has undergone additional quality reviews and received further record additions. On the one hand, we detected that we had erroneously included in the dataset around 150,000 records of *Olea europaea* that corresponded to orchards of cultivated material and not to wild populations of *Olea europaea* subsp. *sylvestris*. Therefore, these records were removed from the dataset. On the other hand, we included in the dataset the data corresponding to the six genetic reserves of CWR established

in the Sierra del Rincón Biosphere Reserve in Madrid. We also included a set of records that were manually searched for those species that had ten or fewer populations in the original dataset. A specific procedure was set for each of them to search for additional records of population occurrence using other data sources (internet, bibliography, citizen science data, public herbaria, communication with botanists, etc.) and to find out if the cause behind the low number of records could be due to a narrow natural distribution or, alternatively, to poor recording in the GBIF database of their real distribution. This work started at the end of the previous semester and was completed this semester. A total of 532 occurrence records corresponding to 34 species were retrieved this way and included in the dataset. Table 1 provides an account of the distribution of these records and the species involved.

Table 1. List of CWR/WFP species manually searched for additional occurrence records from those retrieved from GBIF using various sources of information and the number of new records found.

Target species	Number of new records found
<i>Allium ebusitanum</i>	6
<i>Allium pruinatum</i>	2
<i>Argyranthemum haouarytheum</i>	139
<i>Argyranthemum lemsii</i>	31
<i>Argyranthemum lidii</i>	61
<i>Argyranthemum sundingii</i>	25
<i>Argyranthemum tenerifae</i>	43
<i>Argyranthemum winteri</i>	16
<i>Asparagus macrorrhizus</i>	30
<i>Avena murphyi</i>	4
<i>Cynara algarbiensis</i>	4
<i>Dianthus rupicola</i>	12
<i>Lens ervoides</i>	2
<i>Limonium benmageci</i>	2
<i>Limonium macrophyllum</i>	40
<i>Limonium perezii</i>	7
<i>Limonium redivivum</i>	15
<i>Limonium relicticum</i>	3
<i>Limonium spectabile</i>	5
<i>Limonium vigaroense</i>	8
<i>Medicago citrina</i>	10
<i>Patellifolia webbiana</i>	20
<i>Rubus brigantinus</i>	3
<i>Rubus galloecicus</i>	2
<i>Rubus muricola</i>	1
<i>Rubus pauanus</i>	1
<i>Rubus peratticus</i>	3
<i>Rubus sampaioanus</i>	1
<i>Sideritis osteoxylla</i>	1
<i>Solanum lidii</i>	23
<i>Thymus webbianus</i>	2
<i>Trifolium boissieri</i>	1
<i>Trifolium lucanicum</i>	7
<i>Trigonella anguina</i>	2

As a result of all these changes, the dataset currently contains 1,806,712 records corresponding to 611,612 populations (with a unique POPID identifying them), that covers all 521 species. The median number of populations is 290 and 25% have less than 70 populations, whereas another 25% of the species has more than 1,136 populations. With this work we comply with the deliverable ‘Identify populations from the CWR priority list of the National Strategy for CWR conservation’, which was due in this semester (31/03/2023).

Figure 1 summarizes the flow of information and the curation steps followed to obtain the CWR/WFP population distribution dataset.

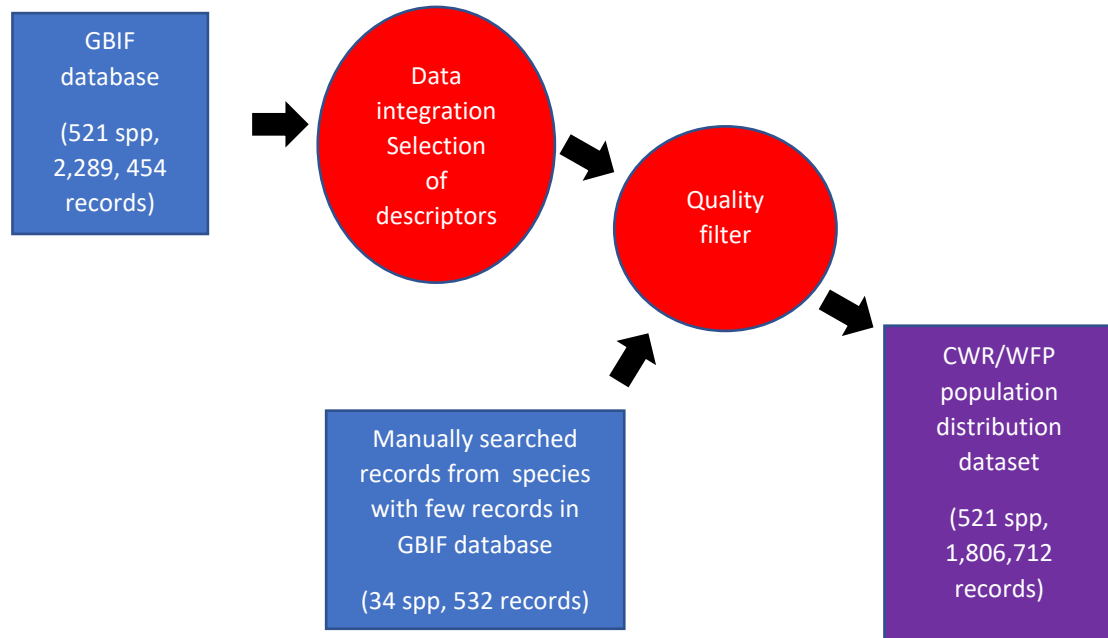


Figure 1. Flow of information and curation steps followed to obtain the CWR/WFP population distribution dataset.

Identify key public and private institutions for the organization of a national network of data providers

The purpose of this directory is to build a national network of contacts from the different public administrations and private organizations that could, on the one hand, contribute to the verification of the distribution data obtained through searches in the most important national and international databases, and on the other hand, facilitate the establishment of genetic reserves.

The directory was compiled in the previous semester since this deliverable had to be finalized by 31/12/2022. It contains 172 contacts and includes representatives from both the Ministry of Agriculture and of Ecological Transition, National Parks Autonomous Agency, wildlife, agriculture, livestock trails and environmental education departments of the 17 autonomous communities, also from national biodiversity databases, and NGO's and Foundations related to organic agriculture, ecosystem restoration and land custody, amongst others. The resulting database can be found [here](#).

During this period, we have continued to approach the main public and private institutions and organizations, contacting them by telephone and/or email, to ask for their willingness to

collaborate and take part in the conservation and use of CWR and WFP. As a result, 51% of the contacts have now explicitly confirmed their willingness to collaborate.

Prepare a draft of the national database structure

Our working team is committed to having the Spanish national *in situ* database built in compliance with the 'FAIR principles' of findable, accessible, interoperable, and reusable data, and has proposed a data flow scheme that is represented in Figures 2 and 3. At the beginning of the process, the data coming from the different sources needs to be integrated into a common structure and handled to filter out the records of insufficient quality. This leads to the generation of the CWR distribution database. Using this database in conjunction with the protected area network layer and the ecogeographic land characterization map (ELC map) as a proxy for identifying populations with different genetic adaptations within species, we will use richness and complementarity analyses to select the most appropriate wild populations for the establishment of genetic reserves (Figure 2). Both the most appropriate wild populations identified and the populations of threatened CWR that are already conserved *in situ* —because of the implementation of recovery plans for threatened species— are candidate populations for the establishment of genetic reserves through a top-down approach (*i.e.*, the initiative is taken by national or autonomic administration and legislation). Additional candidate populations for the establishment of genetic reserves can arise from a bottom-up approach involving farmers or, municipalities at a local scale. They may also arise from initiatives taken by specific interest groups linked to a particular crop (Figure 3).

During the course of this reporting period, we have completed a draft of the data management plan of the Spanish Inventory of *In Situ* Crop Wild Relatives and Wild Food Plants, which can be accessed [here](#). This data management plan contains information about the objectives of the project, the type and formats of data collected, the types and formats of data generated, the origin of data, the expected size of data, the provisions made to approach de FAIR principles, allocation of resources, data security and ethical aspects. This draft data management plan will be presented to the competent authorities leading the National Program on Plant Genetic Resources in the Ministry of Agriculture for further discussion and approval.

Additional work has been undertaken concerning both the selection of descriptors and the preparation of the database structure. After further review, out of the 74 descriptors we selected in the previous semester, we have ended up with 69. These include 38 of the 40 descriptors detailed in Annexes I and II of the document "Principles for the inclusion of CWR Data in EURISCO", leaving out the SUBTAXA and SUBAUTH descriptors. These two have been left out because the Spanish CWR/WFP priority list only includes taxa defined at the species level. At the same time, we have thoroughly worked on the database structure going through several versions until we have reached version 8, which is the one we present at the end of this report in Annex 1. The present database model is structured in three major subject areas, namely 'Taxon Information', 'Population Information' and 'Observation Records'. These subject areas are then composed of several entities in which the different attributes/descriptors are included.

With this work we comply with the deliverable 'Prepare a draft of the national database structure', which is due on 30/09/2023.

We are currently testing the implementation of the database in MongoDB, a NoSQL, schemaless, document database used to build highly scalable internet applications that can work with the json text-based data exchange format. This platform provides high performance in the query operations to select the data necessary to integrate into other data sources such as the National Inventory.

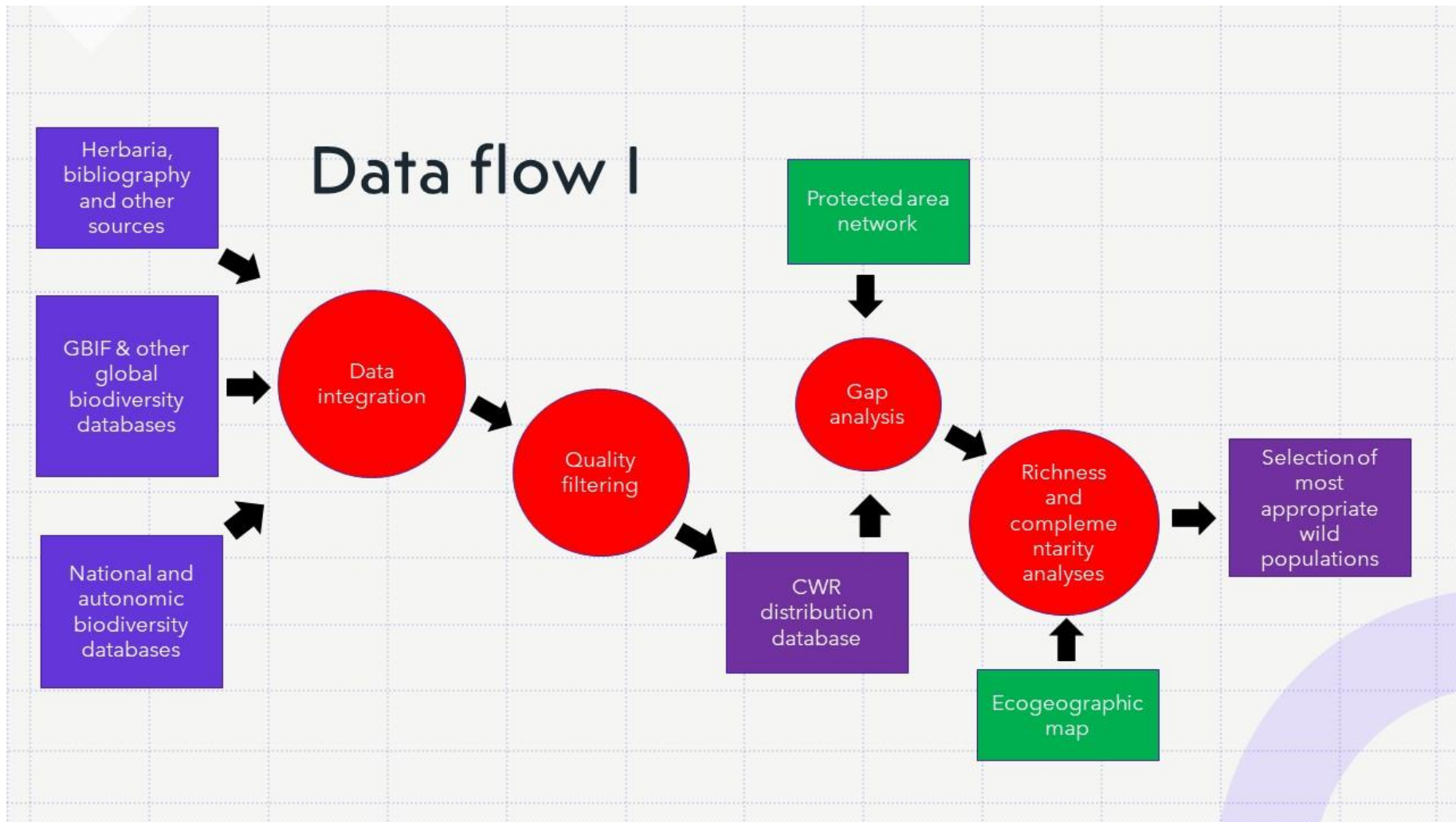


Figure 2. Data flow from sources of information to the generation of the CWR distribution database and the selection of most appropriate wild populations for the establishment of genetic reserves for in situ conservation

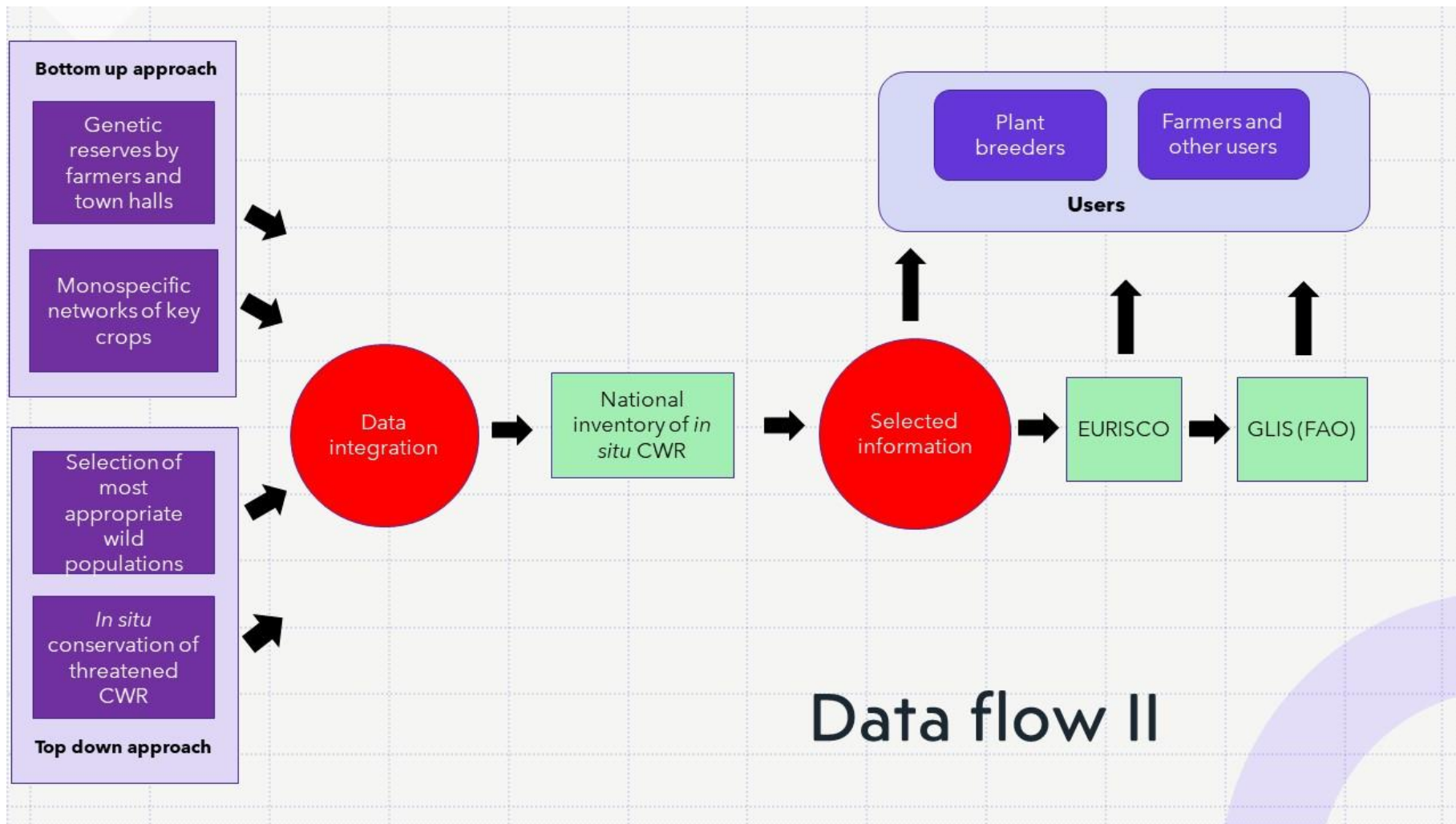


Figure 3. Data flow from different approaches when establishing genetic reserves for the creation of the National Inventory of *in situ* CWR. Subsequently, selected information from the National Inventory is made available to the users directly through national portals or indirectly through EURISCO and GLIS.

Collect and organize available data according to the agreed principles and data exchange format

This semester we have completed preliminary data for most descriptors at taxon level and have also reviewed and updated the contents of the RELATEDCROP, GENEPOOL and LEGSTATUS descriptors. It is worth noting that the current knowledge about the crossability between CWR species and their crops is still low. The most updated source of information available is [GRIN-Global](#) although we have also used the [inventory of Harlan & de Wet](#) because of its simplicity. However, it lacks information about the CWR of many genera, especially those that are not related to human food and forage and fodder crops. In our database, 55% of the priority CWR/WFP species do not have information associated to the GENEPOOL descriptor. We have performed bibliographic searches for some genera to obtain additional genepool information to what is available in GRIN-Global.

Some of the population-level data has been completed with the information obtained from GBIF. In addition, we have fed population-level data with information obtained from thematic layers based on the geographic coordinates of the populations. This includes the descriptors ADM1 and ADM 2, concerning administrative units, ELEVATION, BIOGEOGRAPHIC, ECOGEOGRAPHIC and SITEPROT. Further details about the definition of these descriptors are presented in Annex 1.

To select the ‘most appropriate wild populations’ (*sensu* Maxted et al., 2015), *i.e.* candidate locations to promote the establishment of genetic reserves with the collaboration of the departments of agriculture and environment from the autonomic public administrations, we first generated an ecogeographical land characterization map (herein, ELC) of Spain to characterize the territory in different ecogeographical units. Assuming that the heterogeneity in environmental pressures shapes the distribution patterns of adaptive genetic diversity among the populations of each species, the different ecogeographical units in which a species occur constitute a proxy to the adaptive genetic diversity existing among populations. We used the “ELCmaps” tool of CAPFITOGEN3 (Parra Quijano et al., 2021) to generate an ELC map which contained 27 different ecogeographical units. The next step was to exclude from the analysis those species that currently have legal protection under European, national, or subnational regulations. These regulations establish an obligation to monitor *in situ* their populations and, for the highest protection categories, also the implementation of recovery plans including *in situ* conservation actions. With the remaining species, *i.e.* those that are not legally protected, we conducted a gap analysis by combining layers of both protected natural areas and Natura 2000 sites in Spain to finally select those populations that occur in protected areas. These populations are more suitable for long-term *in situ* conservation and less vulnerable to human interventions. Considering each combination of “target species – ecogeographical unit” as the conservation target, we then conducted a complementarity analysis using the “Complementa” tool of CAPFITOGEN3. We used 1x1km cells as the basic unit to carry out the complementarity analysis. The first 50 1x1km cells in the ranking provide the highest number of “target species – ecogeographical unit” combinations in the lowest number of sites. Therefore, they were tentatively selected as the candidate sites to establish genetic reserves in Spain. These 50 tentative candidate sites would conserve 3,967 populations, 2,878 target “species – Agreement no. L21ROM198 – “3rd Interim Technical Report [01 Jan 2023 – 30 Jun 2023] Page 13

ecogeographical unit” combinations and 321 CWR/WFP species. This represents 45% of the target “species – ecogeographical unit” combinations and 82% of the target species. We are currently exploring the development and application of alternative ecogeographical maps that would structure the territory in a lower number of ecogeographical units. We are also exploring different species-based or “species – ecogeographical unit” combination-based approaches in the complementarity analysis to find out which of them provides optimal results.

We had previously identified populations of threatened priority CWR/WFP which are legally protected and should, in theory, have a management plan where *in situ* monitoring and other active conservation activities are implemented. During the past semester, we searched for these management plans and found out that for some of the protected species there is no such management plan approved or implemented. We have identified just 12 species that have a documented recovery or conservation plan and 27 additional species are included in the management plans of certain zones of special conservation (ZSC) of the Natura 2000 network.

The records of the populations of the [genetic reserves established](#), ‘most appropriate CWR/WFP populations’, and legally protected species that are actively *in situ* conserved still need to be completed with information corresponding to the NICODE, MNGINSTCODE, PUID, SAMPSTAT, POPSRC, CONSACTION, MLSSTAT, MNGINSTNAME, LIAISONCODE, LIAISONNAME, OTHERNUMB, HERBCODE, HERBNAME, SPECNUMB, and LINKS descriptors.

Finally, we have implemented a database version control that tracks every change made to the database and conserves the different versions.

Provide the data to EURISCO

We aim at following the guidelines provided in the document ECPGR “[Principles for the Inclusion of CWR Data in EURISCO](#)”. We have started a discussion with the Ministry of Agriculture on the selection of the records to be sent to EURISCO, including the assessment on whether it is appropriate to upload records of populations of threatened CWR/WFP protected by law.

References

Castroviejo, S. (Coord. Gen.) (1986-2022) Flora Iberica 1-21. Real Jardín Botánico, CSIC, Madrid.

Chamberlain S, Boettiger C (2017) R Python, and Ruby clients for GBIF species occurrence data, PeerJ PrePrints <https://doi.org/10.7287/peerj.preprints.3304v1>.

Iriondo, JM Coord. (2011). Atlas y Libro Rojo de la Flora Vasculare Amenazada de España. Manual de metodología del trabajo corológico y demográfico. Dirección General de Medio Natural y Política Forestal (Ministerio de Medio Ambiente, y Medio Rural y Marino)-Sociedad Española de Biología de la Conservación de Plantas. Madrid, 70 pp.

Maxted N, Avagyan A, Frese L, Iriondo JM, Magos Brehm J, Singer A, Kell SP. (2015) ECPGR Concept for in situ conservation of crop wild relatives in Europe. Wild Species Conservation in Agreement no. L21ROM198 – “3rd Interim Technical Report [01 Jan 2023 – 30 Jun 2023] Page 14

Genetic Reserves Working Group, European Cooperative Programme for Plant Genetic Resources, Rome, Italy.

Parra Quijano, M., Iriondo, J.M., Torres, M.A., López, F., Phillips, J., Kell, S. (2021). CAPFITOGEN3: Una caja de herramientas para la conservación y promoción del uso de la biodiversidad agrícola. Universidad Nacional de Colombia, Sede Bogotá - Facultad de Ciencias Agrarias.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Annex 1. National Inventory of Crop Wild Relatives and Wild Food Plants of Spain (CWR-V8). Logical database model documentation.

Contents

1. Model details	18
2. Entities	20
2.1. Entity TAX_USES.....	20
2.2. Entity RELATEDCROP	20
2.3. Entity POPULATION.....	21
2.4. Entity MANAGEMENT	24
2.5. Entity HERBARIUM	26
2.6. Entity LINKS	27
2.7. Entity SITEDESCRIPTION.....	27
2.8. Entity RECORD.....	29
2.9. Entity TAXON.....	30
2.10. Entity CONSERVATION MEASURES	32
2.11. Entity EX SITU CONSERVATION	33
2.12. Entity INVENTORY	33
3. Entity relationships	37
3.1. Entity relationship HERBARIUM_RECORD	37
3.2. Entity relationship RECORD_SITEDESCRIPTION	37
3.3. Entity relationship RELATEDCROP_TAXON	37
3.4. Entity relationship TAXON_USES	37
3.5. Entity relationship LINKS_POPULATION	37
3.6. Entity relationship RECORD_POPULATION	37
3.7. Entity relationship RECORD_TAXON	38
3.8. Entity relationship EX SITU CONSERVATION_POPULATION	38
3.9. Entity relationship INVENTORY_MANAGEMENT	38
3.10. Entity relationship CONSERVATION MEASURES_INVENTORY	38
4. Areas	39
4.1. Area Taxon Information	39
4.2. Area Population Information	39
4.3. Area Observation records	39

1. Model details

Model name: CWR-V8

Created with Vertabelo.com

The **purpose** of the data collection is to generate a database for the Spanish Inventory of In Situ Conserved Crop Wild Relatives (CWR) and Wild Food Plants (WFP) following the FAIR principles and to transfer a selected part of this information to EURISCO's database (<https://eurisco.ipk-gatersleben.de/apex/f?p=103:1:0::::>) according to the principles stated in the technical report "Principles for the inclusion of CWR data in EURISCO" (<https://www.ecpgr.cgiar.org/resources/ecpgr-publications/publication/principles-for-the-inclusion-of-cwr-data-in-eurisco-2022>). The targeted CWR and WFP are those indicated in Spain's priority list of the National Strategy for Conservation and Use of Crop Wild Relatives and Wild Food Plants (https://www.mapa.gob.es/es/agricultura/temas/medios-de-produccion/mapa_estrategiadeconservacion_04_tcm30-636650.pdf).

The present database model is structured in three major subject areas, namely 'Taxon Information', 'Population Information' and 'Observation Records' (Figure 1). These subject areas are then composed of several entities in which the different attributes are included. The relationships between entities are of the types 'one to one' and 'one to many'.

The purpose of the database is to identify and select a set of CWR/WFP populations that are, either being actively *in situ* conserved or especially valuable and should be considered for *in situ* conservation in genetic reserves by the public administration. The process to obtain this information starts with the download of occurrence records from different global, regional, national and local sources, especially from GBIF. Then, these data are curated and aggregated to the population level, considering that occurrence records that are less than 500m apart from each other could belong to a single population. Additional information is fed to the database through GIS tools by intersecting the geographical coordinates of the populations with thematic layers. On the other hand, relevant information concerning the legal protection status and threat categories of the species has been gathered from other sources. The application of different criteria allows the selection of populations that are included in the National Inventory. Therefore, the Inventory entity in the database model appears as an inheritance of the Population entity, indicating that only a subset of the population records acquires the attributes included in the Inventory entity.

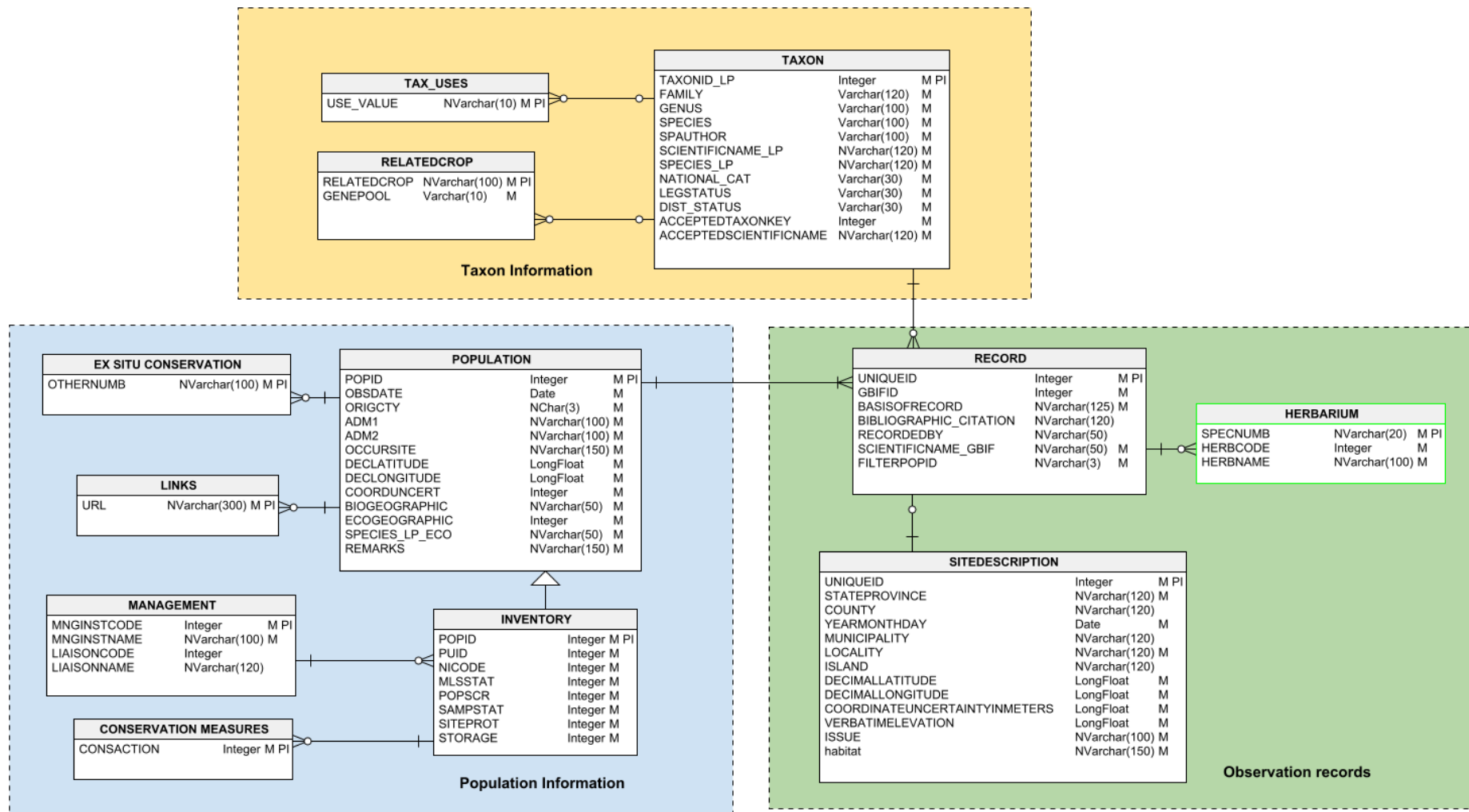


Figure 1. General overview of the database model of the National Inventory of Crops Wild Relatives and Wild Food Plants of Spain

2. Entities

2.1. Entity TAX_USES

Description: What is the plant used for. CWR: crop wild relative; WFP: wild food plant; OTHER: other ecosystem services. Multiple entries are separated by a semicolon (;) without space. NOTE: This descriptor should be used when NIs contain records of wild plants other than crop wild relatives.

2.1.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
USE_VALUE	NVarchar(10)	yes	Yes	CWR: crop wild relative; WFP: wild food plant; OTHER: other ecosystem services

2.2. Entity RELATEDCROP

Description: The scientific name(s) of the crops to which the taxon is related.

2.2.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
RELATEDCROP	NVarchar(100)	yes	Yes	Multiple entries are separated by a semicolon (;) without space. Example: Brassica oleracea L.;Brassica juncea (L.) Czern. NOTE: Depending on the interests of each country, alternative crop classification systems may be used. When possible, use the list of

				crop names used by GRIN.
GENEPOOL	Varchar(10)	yes	No	GP1: primary; GP2: secondary; GP3: tertiary

2.3. Entity POPULATION

Description: This entity includes the geographic and temporal attributes that are related to each known population of the CWR/WFP species of the Catalogue.

2.3.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
POPID	Integer	yes	yes	The identifier (sequential number or code) that the National Inventory uses to identify each population. Each distinct population should be given a population unique identifier. NOTE: This descriptor is mandatory for upload to EURISCO.
OBSDATE	Date	yes	no	The most recent date the population was observed. YYYY: year; MM: month; DD: day. Missing data (MM or DD) should be indicated with '00' [double zero]. NOTE: This descriptor is considered mandatory in the FAO list.
ORIGCTY	NChar(3)	yes	no	Country where the CWR population was observed or inventoried. Use the Three-letter ISO 3166-1 code of the country where the site is located. NOTE: This descriptor is considered mandatory in the FAO list.
ADM1	NVarchar(100)	yes	no	Official name of the administrative level below country level (ORIGCTY) which

				corresponds to the occurrence site. In Spain, it corresponds to Autonomous Community. This value has been extracted for each population record using GIS tools.
ADM2	NVarchar(100)	yes	no	Official name of the administrative level below the ADM1 level. In Spain, it corresponds to Province. This value has been extracted for each population record using GIS tools.
OCCURSITE	NVarchar(150)	yes	no	Location information below the country level that describes the site where the population sample was observed, inventoried or collected, preferably in English. This might include the distance in km and direction from the nearest town, village or map grid reference point, (e.g. 7km south of Curitiba in the state of Parana). NOTE: This descriptor may be used to indicate the extension of the population in the case of ubiquitous widespread taxa.
DECLATITUDE	LongFloat	yes	no	Latitude of the site expressed in decimal degrees. Positive values are north of the Equator; negative values are south of the Equator (e.g. -44.6975). NOTE: Latitude and longitude should be provided using the WGS84 datum, as this is the standard used by GBIF.
DECLONGITUDE	LongFloat	yes	no	Longitude of the site expressed in

				decimal degrees. Positive values are east of Greenwich Meridian; negative values are west of Greenwich Meridian (e.g. -120.9123). NOTE: Latitude and longitude should be provided using the WGS84 datum, as this is the standard used by GBIF.
COORDUNCERT	Integer	yes	no	Uncertainty associated with the coordinates in metres. Leave the value empty if the uncertainty is unknown. NOTE: This descriptor may also be used to indicate the extension of the population in the case of ubiquitous widespread taxa.
BIOGEOGRAPHIC	NVarchar (50)	yes	no	Biogeographic region in which the population occurs: Atlantic (ATL), Mediterranean (MED), Alpine (ALP) and Macaronesian (MAC). Obtained by intersecting the geographic coordinates of the population with a map of the biogeographic regions in a GIS
ECOGEOGRAPHIC	Integer	yes	no	Ecogeographic category of each population. Numeric value given between 1 and 27, a number for each created category.
SPECIES_LP_ECO	NVarchar (50)	yes	no	Indicates the combination between the category Species_LP and the ECOGEOGRAPHIC number given for each population.
REMARKS	NVarchar (150)	yes	no	The Remarks field is used to add notes or to elaborate on descriptors with value 99 or 999 (= Other). Prefix

				remarks with the field name they refer to and a colon (:.) without space (e.g. SITESTAT:riverside). Distinct remarks referring to different fields are separated by semicolons without space.
--	--	--	--	---

2.4. Entity MANAGEMENT

Description: Information about the institutions in charge of managing the site where the wild population occurs

2.4.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
MNGINSTCODE	Integer	yes	yes	FAO WIEWS code of the institution responsible for the population (e.g. protected area authority, nature reserve manager, national park manager, private landowner, etc.). NOTE: This descriptor is new and mandatory for upload to EURISCO.
MNGINSTNAME	NVarchar(100)	yes	no	Name and short address of the institution responsible for the population, to be used if the previous file is

				empty. NOTE: the description was changed from the FAO list.
LIAISONCODE	Integer	no	no	FAO WIEWS code of the institution that can liaise between the organization managing the CWR population and the interested user to facilitate access to the material. NOTE: If this and the next field are empty, the managing organization can be approached for access.
LIAISONNAME	NVarchar (120)	no	no	Name, and brief address, of the institution that can liaise between the organization managing the CWR population and the interested user. NOTE: If this and the previous field are empty, the managing institute can be approached for access.

2.4.2. Entity identifiers

Entity identifier name	Attributes	Description
MANAGEMENT_ak_1	LIAISONCODE	

2.5. Entity HERBARIUM

Description: This entity holds information about the herbarium specimens related to the observation records held in public herbaria

2.5.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
SPECNUMB	NVarchar (20)	yes	Yes	This is the unique identifier for herbarium specimens collected and is assigned when a specimen is entered into the collection. NA values allowed
HERBCODE	Integer	yes	No	Index Herbariorum code of the herbarium that holds a herbarium specimen of the population. NA values allowed.
HERBNAME	NVarchar (100)	yes	No	Name of the herbarium that holds herbarium specimens of the population. This descriptor should only be used if HERBCODE is empty.

2.5.2. Entity identifiers

Entity identifier name	Attributes	Description
HERBARIUM_ak_1	HERBCODE	

2.6. Entity LINKS

Description: Links where additional information about the CWR population can be found

2.6.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
URL	NVarchar(300)	yes	Yes	One or more URLs where further information about the CWR can be found. Multiple values are separated by a semicolon without space. NOTE: Only use one URL where the associated information remains updated.

2.7. Entity SITEDESCRIPTION

Description: Site description where CWR population was observed or inventoried

2.7.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
UNIQUEID	Integer	yes	yes	Unique record identifier. Provided by the research team maintaining the database.
STATEPROVINCE	NVarchar(120)	yes	no	The name of the next smaller administrative region than country (state, province, canton, department, region, etc.) in which the Location occurs. For detailed information and examples, please

				visit: https://dwc.tdwg.org/terms/ . NA values allowed.
COUNTY	NVarChar(120)	no	no	The full, unabbreviated name of the next smaller administrative region than stateProvince (county, shire, department, etc.) in which the Location occurs. For detailed information and examples, please visit: https://dwc.tdwg.org/terms/ . NA values allowed.
YEARMONTHDAY	Date	yes	no	Date of the observation formed by adding year, month and day fields of GBIF
MUNICIPALITY	NVarChar(120)	no	no	The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs. Do not use this term for a nearby named place that does not contain the actual location. For detailed information and examples, please visit: https://dwc.tdwg.org/terms/ . NA values allowed.
LOCALITY	NVarChar(120)	yes	no	The specific description of the place. For detailed information and examples, please visit: https://dwc.tdwg.org/terms/ . NA values allowed.
ISLAND	NVarChar(120)	no	no	The name of the island on or near which the Location occurs. For detailed information and examples, please visit: https://dwc.tdwg.org/terms/ . NA values allowed.
DECIMALLATITUDE	LongFloat	yes	no	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive. NA values not allowed. For detailed information, please visit: https://dwc.tdwg.org/terms/
DECIMALLONGITUDE	LongFloat	yes	no	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic center of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.
COORDINATEUNCERTAINTYINMETERS	LongFloat	yes	no	The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the

				value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term.
VERBATIMELEVATION	LongFloat	yes	no	The original description of the elevation (altitude, usually above sea level) of the Location. For further information on this field, please visit: https://dwc.tdwg.org/terms/#dwc:coordinateUncertaintyInMeters
ISSUE	NVarchar(100)	yes	no	Any issue found during processing and interpretation or the record. See https://gbif.github.io/gbif-api/apidocs/org/gbif/api/vocabulary/OccurrenceIssue.html for possible values.
habitat	NVarchar(150)	yes	no	A category or description of the habitat in which the event occurred. NA values allowed. For further information, please visit: https://dwc.tdwg.org/terms/

2.8. Entity RECORD

Description: Observation records

2.8.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
UNIQUEID	Integer	yes	yes	Unique record identifier. Provided by the research team maintaining the database.
GBIFID	Integer	yes	no	Numerical, stable identifier assigned by GBIF to an Occurrence record. For detailed information, please visit: https://gbif.github.io/dwc-api/apidocs/org/gbif/dwc/terms/GbifTerm.html
BASISOFRECORD	NVarchar(125)	yes	no	The specific nature of the data record. This could be a human observation, a preserved specimen, among others. For further information, please visit: https://dwc.tdwg.org/terms/ If BASISOFRECORD=herbarium specimen then HERBARIUM link exists;
BIBLIOGRAPHIC_CITATION	NVarchar(120)	no	no	A bibliographic reference for the resource as a statement indicating how this record should be cited (attributed) when used. From Dublin Core,

				"Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible." The intended usage of this term in Darwin Core is to provide the preferred way to cite the resource itself - "how to cite this record". Note that the intended usage of dcterms:references in Darwin Core, by contrast, is to point to the definitive source representation of the resource - "where to find the as-close-to-original reference, if one is available.
RECORDED BY	NVarchar (50)	no	no	A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first.
SCIENTIFICNAME_GBIF	NVarchar (50)	yes	no	Scientific name with author according to source providing data to GBIF (which may differ to acceptedScientificName_GBIF)
FILTERPOPID	NVarchar (3)	yes	no	Field that indicates if the record is selected as representative of the population. Yes: when the record provides the coordinates for the population. No: otherwise

2.8.2. Entity identifiers

Entity identifier name	Attributes	Description
RECORD_ak_1	GBIFID	

2.9. Entity TAXON

Description: This entity provides taxonomic and nomenclatural information about the target species. It also provides information concerning its threat status and legal protection status.

2.9.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description

			fier	
TAXONID_LP	Integer	yes	Yes	Unique taxon identifier according to the one given in the "Standard List (=Lista Patrón)" of Spain (hereafter LP). The Standard List is a monitored, structured, normalized and agreed list, basic for the proper operation of the Integrated System of Information for the Spanish Inventory of Natural and Biodiversity Heritage. With this list, a given element (i.e. plant, animal, fungus) is named in the same way by all organisms and entities involved in its management. The list is available in: https://www.miteco.gob.es/en/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/BDN_listas_patron.aspx . Please, click on "Lista patrón de las especies silvestres presentes en España" for download. NA values not allowed.
FAMILY	Varchar (120)	yes	no	Taxon family, in Latin. Initial uppercase required. NA values not allowed.
GENUS	Varchar (100)	yes	no	Genus name for the taxon. If applied taxonomies among sources differ, Taxonomy in the Standard List prevails. NA values not allowed.
SPECIES	Varchar (100)	yes	no	Species epithet portion of the scientific name. If applied taxonomies among sources differ, Taxonomy in the Standard List prevails. NA values not allowed.
SPAUTHOR	Varchar (100)	yes	no	Authority for the species name. NA values not allowed.
SCIENTIFICNAME_LP	NVarchar (120)	yes	no	Scientific name with author according to the LP. NA values not allowed.
SPECIES_LP	NVarchar (120)	yes	no	Scientific name without author according to the LP. NA values not allowed.
NATIONAL_CAT	Varchar (30)	yes	no	The Red List category according to national criteria. CR: Critically endangered; EN: Endangered; VU: Vulnerable; NT: Near threatened; LC: Least concern; DD: Data deficient; NE: Not evaluated. NOTE: Use the most recent assessment. NA values not allowed.
LEGSTATUS	Varchar (30)	yes	no	Informs whether the taxon is legally protected at the European, national or subnational levels. HD: Habitats Directive; NAT: National level; SUB: Subnational level.
DIST_STATUS	Varchar (30)	yes	no	The distribution status of the taxon within the geographic area of the checklist or inventory, indicating whether it is a National endemic,

				Regional endemic, Cosmopolitan, or Unknown. NOTE: 'Regional' is defined here as a geographic area comprising different countries rather than a sub-unit within a country.
ACCEPTEDTAXONKEY	Integer	yes	no	The GBIF backbone key of the accepted taxon key (taxonkey corresponding to acceptedScientificName GBIF).
ACCEPTEDSCIENTIFICNAME	NVarchar (120)	yes	no	Accepted scientific name, including authority, according to GBIF backbone. For more information, please visit:

2.9.2. Entity identifiers

Entity identifier name	Attributes	Description
TAXON_ak_1	ACCEPTEDTAXONKEY	

2.10. Entity CONSERVATION MEASURES

Description: This entity informs about the different conservation measures that are put in place regarding the target population

2.10.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
CONSACTION	Integer	Yes	yes	Indication whether conservation actions related to the population are in place. Use the IUCN classification scheme for conservation actions in place. 0: No conservation actions 1: Monitoring and Planning 2: Land/Water Protection and Management 3: Species Management 4: Education and Legislation 99: Other (elaborate in REMARKS field) NOTE: There

				can be more than one answer. For example: 1; 3.
--	--	--	--	---

2.11. Entity EX SITU CONSERVATION

Description: This entity informs about the seed accessions corresponding to the target populations that are preserved in genebanks.

2.11.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
OTHERNUMB	NVarchar(100)	yes	yes	Any other identifiers known to exist in ex situ collections for this population. INSTCODE and identifier are separated by a colon without space. Pairs of INSTCODE and identifier are separated by a semicolon without space. Example: INSTCODE:identifier;INSTCODE:identifier;... When the institute is not known, the identifier should be preceded by a colon.

2.12. Entity INVENTORY

Description: This entity contains the attributes that are only given to the populations that are selected to be part of the Spanish Inventory of CWR and WFP.

2.12.1. Attributes

Attribute name	Type	Is mandatory	Is primary identifier	Description
POPID	Integer	yes	yes	The identifier (sequential number or code) that the National Inventory uses to identify each population. Each distinct population should be given a population unique identifier. NOTE: This descriptor is mandatory for upload to EURISCO. NA values not allowed.

PUID	Integer	yes	no	Any persistent, unique identifier (preferably a DOI) assigned to the population so it can be unambiguously referenced at the global level and the information associated with it harvested through automated means. Report only one PUID for each population. The Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture (PGRFA) is facilitating the assignment of a persistent unique identifier (PUID), in the form of a DOI, to PGRFA at the accession level. NOTE: This descriptor should be assigned only to those CWR populations that the National Focal Point considers as long-term available sources of germplasm (e.g. the population is being monitored and is potentially available under the terms of the MLS).
NICODE	Integer	yes	no	Code identifying the National Inventory; the Three-letter ISO 3166-1 code of the country preparing the National Inventory. Exceptions are possible if agreed with EURISCO, such as NGB. Example: NLD
MLSSTAT	Integer	yes	no	The status of the in situ accession of the CWR population with

				<p>regards to the Multilateral System of Access and Benefit-sharing of the International Treaty, if available. 0: Not available under the MLS. 1: Available under the MLS.</p> <p>NOTE: The definition of this descriptor in Alercia et al. (2021) refers to the corresponding ex situ accession, but in the CWR-NI it should be applied to the in situ accession. The ex situ accession will have its own MLSSTAT descriptor in the genebank inventories.</p>
POPSCR	Integer	yes	no	<p>Habitat of the occurrence site of the population(s). The coding scheme can be applied either by using the general codes or the more specific codes. Multiple values are separated by a semicolon without space. 20: Farm or cultivated area (21: Field, 22: Orchard, 23: Backyard, kitchen or home garden, 24: Fallow land, 25: Pasture, 28: Park). 10: Wild (11: Forest or woodland, 12: Shrubland, 13: Grassland, 14: Desert or tundra, 15: Aquatic habitat) 60: Weedy, disturbed or ruderal habitat (61: Roadside, 62: Field margin) 99: Other (elaborate in REMARKS field)</p>
SAMPSTAT	Integer	yes	no	<p>The coding scheme proposed can be</p>

				<p>applied by using the general codes or the more specific codes. 100: Wild (110: Natural, 120: Semi-natural/wild, 130: Semi-natural/sown). 200: Weedy. 999: Other (elaborate in REMARKS field). NOTE: Use 100 for wild populations, unless there is specific information indicating its semi-natural status.</p>
SITEPROT	Integer	yes	no	<p>Indicate whether the site is protected under any legal or official protection. 0: not protected; 1: strict nature reserve; 2: wilderness area; 3: national park; 4: natural monument or feature; 5: habitat/species management area; 6: protected landscape/seascape; 7: protected area with sustainable use of natural resources. 8: other effective conservation measures</p>
STORAGE	Integer	yes	no	<p>Type of germplasm storage. For in situ CWR populations this descriptor should always have the value 60.</p>

2.12.2. Entity identifiers

Entity identifier name	Attributes	Description
INVENTORY_ak_1	PUID	

3. Entity relationships

3.1. Entity relationship HERBARIUM_RECORD

RECORD	1..*	HERBARIUM
	<->	

3.2. Entity relationship RECORD_SITEDESCRIPTION

SITEDESCRIPTION	1..1	RECORD
	<->	

3.3. Entity relationship RELATEDCROP_TAXON

TAXON	1..*	RELATEDCROP
	<->	

3.4. Entity relationship TAXON_USES

TAX_USES	*..1	TAXON
	<->	

3.5. Entity relationship LINKS_POPULATION

POPULATION	1..*	LINKS
	<->	

3.6. Entity relationship RECORD_POPULATION

POPULATION	1..*	RECORD
	<->	

3.7. Entity relationship RECORD_TAXON

TAXON	1..*	RECORD
	<->	

3.8. Entity relationship EX SITU CONSERVATION_POPULATION

POPULATION	1..*	EX SITU CONSERVATION
	<->	

3.9. Entity relationship INVENTORY_MANAGEMENT

MANAGEMENT	1..*	INVENTORY
	<->	

3.10. Entity relationship CONSERVATION MEASURES_INVENTORY

INVENTORY	1..*	CONSERVATION MEASURES
	<->	

4. Areas

4.1. Area Taxon Information

4.1.1. Entities

- TAX_USES
- RELATEDCROP
- TAXON

4.1.2. Entity relationships

- RELATEDCROP_TAXON
- TAXON_USES

4.2. Area Population Information

4.2.1. Entities

- POPULATION
- MANAGEMENT
- LINKS
- CONSERVATION MEASURES
- EX SITU CONSERVATION
- INVENTORY

4.2.2. Entity relationships

- LINKS_POPULATION
- EX SITU CONSERVATION_POPULATION
- INVENTORY_MANAGEMENT
- CONSERVATION MEASURES_INVENTORY

4.3. Area Observation records

4.3.1. Entities

- HERBARIUM
- SITEDESCRIPTION
- RECORD

4.3.2. Entity relationships

- HERBARIUM_RECORD
- RECORD_SITEDESCRIPTION