Stephan Weise

EURISCO extension for *in situ* CWR data

EURISCO training workshop 2023
12–14 September 2023, Plovdiv, Bulgaria

Background

- *In situ* CWR populations are potentially valuable resources for both science and breeding

  →Need to be conserved

  →Need to be made available

- However:
  - Conservation of and access to CWR populations varies significantly
    - Nature conservation organisations
    - Farmer‘s fields
    - Roadsides
    - Not managed at all
  - Information about CWR populations often not available at all

Background

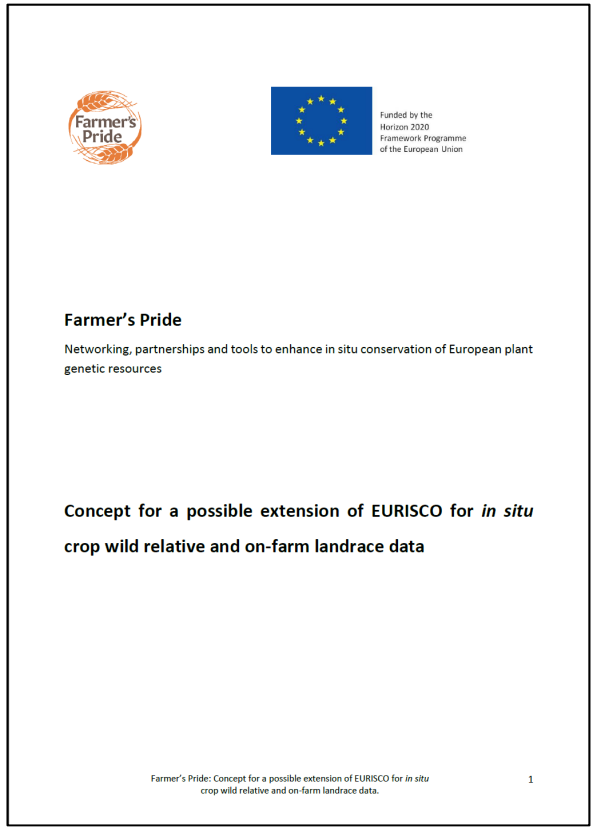
- Currently only *ex situ* data in EURISCO
- *In situ* extension desired for a long time
- Previous activities of the *in situ* community
  - Checklists and descriptors for *in situ* CWR conservation
  - Descriptors for *in situ* LR inventories
  - ECPGR concepts for *in situ* CWR and on-farm conservation
  - …

Background

- Technical concept for extension of EURISCO for *in situ* data developed in Farmer's Pride project (*in situ* CWR + on-farm LR)

- Necessary prerequisite: exchange and regular update of data
  - Rather uncritical for *in situ* CWR
  - Large logistical effort for on-farm LR

  →For the time being, focus on *in situ* CWR data


Farmer's Pride

Funded by the Horizon 2020 Framework Programme of the European Union

Farmer's Pride

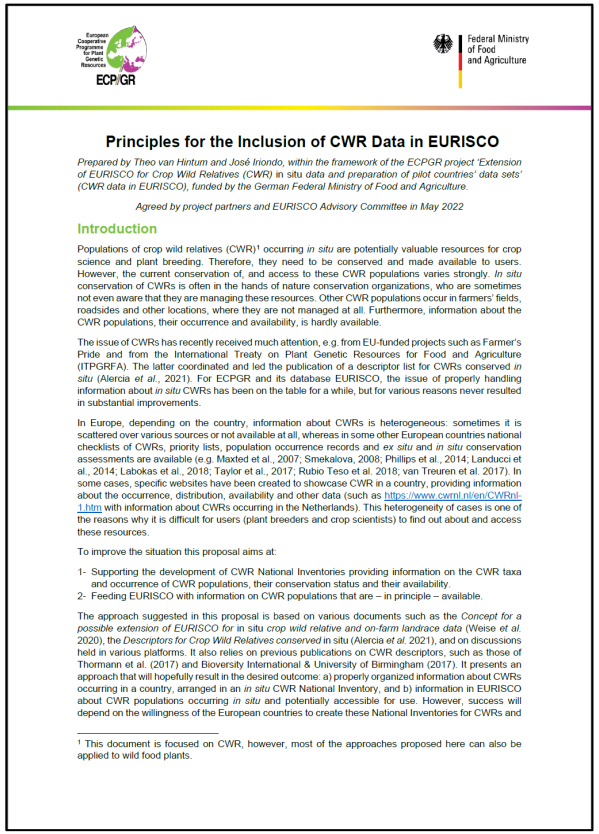
Networking, partnerships and tools to enhance in situ conservation of European plant genetic resources

Concept for a possible extension of EURISCO for *in situ* crop wild relative and on-farm landrace data

Farmer's Pride: Concept for a possible extension of EURISCO for *in situ* crop wild relative and on-farm landrace data. 1

Background

- ECPGR project “Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and preparation of pilot countries’ data sets” (CWR data in EURISCO)
  - Funded by the German Federal Ministry of Food and Agriculture
  - Definition of principles of data inclusion in EURISCO (T. van Hintum/J. Iriondo) → approved by project partners + EURISCO AC
  - Implementation of data integration + extension of public web interface → 10/2022 – 12/2023
  - Pilot countries to provide their data and test the system



ECP/GR

Federal Ministry of Food and Agriculture

Principles for the Inclusion of CWR Data in EURISCO

Prepared by Theo van Hintum and José Iriondo, within the framework of the ECPGR project 'Extension of EURISCO for Crop Wild Relatives (CWR) in situ data and preparation of pilot countries' data sets' (CWR data in EURISCO), funded by the German Federal Ministry of Food and Agriculture.

Agreed by project partners and EURISCO Advisory Committee in May 2022

Introduction

Populations of crop wild relatives (CWR)[1] occurring *in situ* are potentially valuable resources for crop science and plant breeding. Therefore, they need to be conserved and made available to users. However, the current conservation of, and access to these CWR populations varies strongly. *In situ* conservation of CWRs is often in the hands of nature conservation organizations, who are sometimes not even aware that they are managing these resources. Other CWR populations occur in farmers' fields, roadsides and other locations, where they are not managed at all. Furthermore, information about the CWR populations, their occurrence and availability, is hardly available.

The issue of CWRs has recently received much attention, e.g. from EU-funded projects such as Farmer's Pride and from the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). The latter coordinated and led the publication of a descriptor list for CWRs conserved *in situ* (Alercia *et al.*, 2021). For ECPGR and its database EURISCO, the issue of properly handling information about *in situ* CWRs has been on the table for a while, but for various reasons never resulted in substantial improvements.

In Europe, depending on the country, information about CWRs is heterogeneous: sometimes it is scattered over various sources or not available at all, whereas in some other European countries national checklists of CWRs, priority lists, population occurrence records and *ex situ* and *in situ* conservation assessments are available (e.g. Maxted et al., 2007; Smekalova, 2008; Phillips et al., 2014; Landucci et al., 2014; Labokas et al., 2018; Taylor et al., 2017; Rubio Teso et al. 2018; van Treuren et al. 2017). In some cases, specific websites have been created to showcase CWR in a country, providing information about the occurrence, distribution, availability and other data (such as https://www.cwrnl.nl/en/CWRnl-1.htm with information about CWRs occurring in the Netherlands). This heterogeneity of cases is one of the reasons why it is difficult for users (plant breeders and crop scientists) to find out about and access these resources.

To improve the situation this proposal aims at:

1- Supporting the development of CWR National Inventories providing information on the CWR taxa and occurrence of CWR populations, their conservation status and their availability.
2- Feeding EURISCO with information on CWR populations that are – in principle – available.

The approach suggested in this proposal is based on various documents such as the *Concept for a possible extension of EURISCO for* in situ *crop wild relative and on-farm landrace data* (Weise *et al.* 2020), the *Descriptors for Crop Wild Relatives conserved* in situ (Alercia *et al.* 2021), and on discussions held in various platforms. It also relies on previous publications on CWR descriptors, such as those of Thormann et al. (2017) and Bioversity International & University of Birmingham (2017). It presents an approach that will hopefully result in the desired outcome: a) properly organized information about CWRs occurring in a country, arranged in an *in situ* CWR National Inventory, and b) information in EURISCO about CWR populations occurring *in situ* and potentially accessible for use. However, success will depend on the willingness of the European countries to create these National Inventories for CWRs and

[1] This document is focused on CWR, however, most of the approaches proposed here can also be applied to wild food plants.

Approach

- Inclusion of *in situ* CWR data in EURISCO
  - Development of National Inventories for CWR (CWR-NI)
  - Organisation of data flow from CWR-NI to EURISCO

Maintenance site *in situ* CWR

Maintenance site *in situ* CWR

Maintenance site *in situ* CWR

Liaison institute

In situ National Inventory

Maintenance site *in situ* CWR

Maintenance site *in situ* CWR

In situ National Inventory

Maintenance site *in situ* CWR

Liaison institute

In situ National Inventory

...

...

...

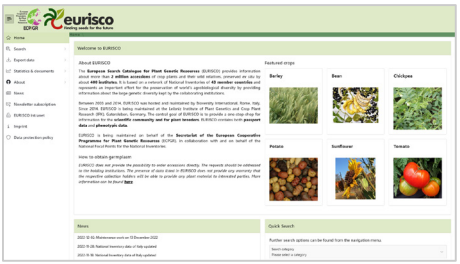
eurisco Finding seeds for the future

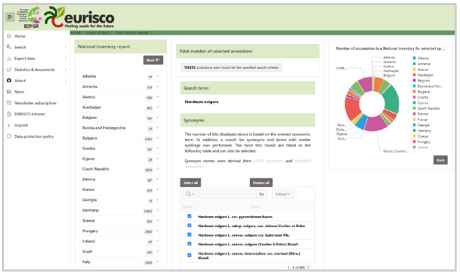
EURISCO-CWR staging area

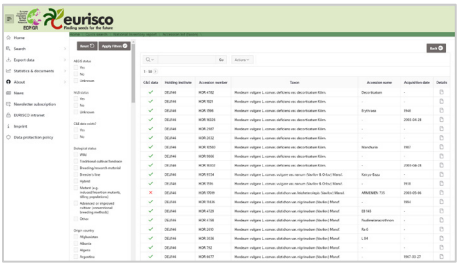
Data integrity checks

Update of EURISCO-CWR dataset

eurisco Finding seeds for the future







National Inventories for CWR

- Identification of CWRs of interest/development of checklist of priority crops
- Development of database structure for information about management and use of CWR-NI
  - In principle up to each country
  - Recommendations of descriptors for the generation of CWR-NI
    - Information at taxon level
      - Used for the generation of the checklist
      - Taxonomy information
      - Crossability of the CWR (genepool)
      - Threat status, protection status, …
      - Related crop
    - Information at population level
      - Descriptors of population site (coordinates, country, habitat, …)
      - Population descriptors (most recent observation date, holding institution, availability, *ex situ* available? herbarium specimen?)
      - Population management descriptors (threats, conservation actions, …)

Data flow from CWR-NI to EURISCO

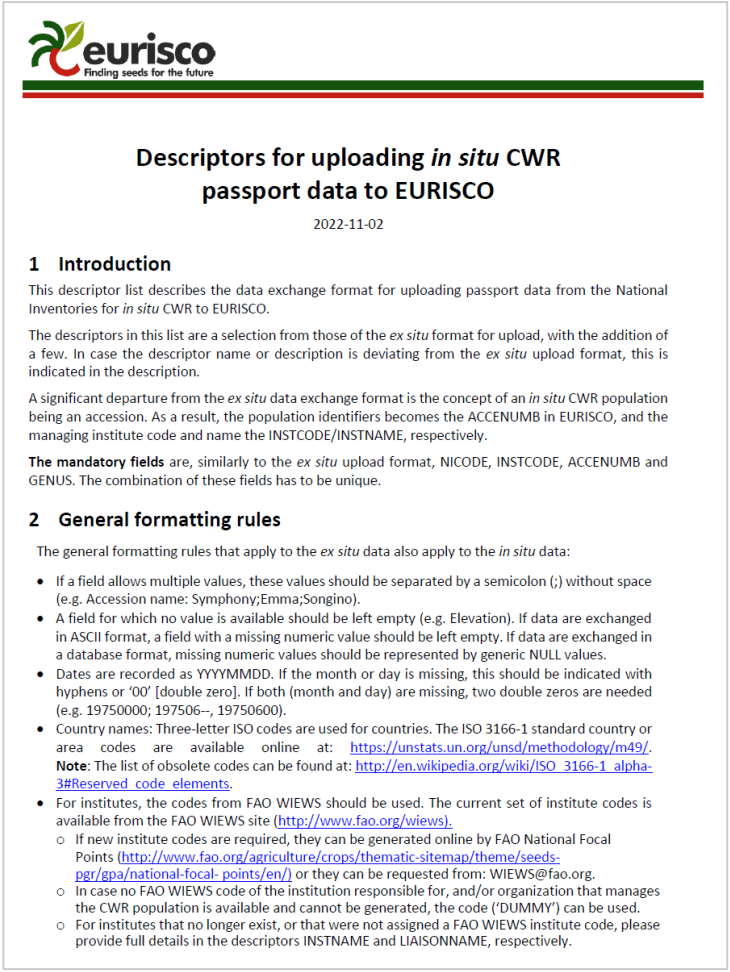
- Actively conserved populations
  - Focus on populations that can be made available to users in principle
  - Probably those that are “actively conserved”
    - Likely to exist
    - Location is known
    - Management institution/person that can be approached to access the material
- Access to *in situ* material
  - There should be a designated pathway to approach managing/liaison institute
    - Get information about how to obtain material
    - Get information about terms and conditions
- What information can be shared with EURISCO?
  - Local decision
    - Which populations should be made visible?
    - Which data should be shared?

Steps to integrate data into EURISCO

- Descriptors for uploading *in situ* CWR passport data
- Upload to EURISCO
  - By an authorised CWR Focal Point (can be the *ex situ* NFP or another)
  - *In situ* CWR data separated from *ex situ* data
- Database schema extension
- Development of import tool for *in situ* CWR data
- Development of procedures for data integrity checks and data integration
- Web interface extension according to user requirements

Recent developments

- Compilation of EURISCO *in situ* CWR passport data standard
  - Mapping of CWR information on current EURISCO structure
  - CWR population considered similar to an *ex situ* accession
    - Population ID like the accession number
    - New concept: institute for liaison between potential user and managing organisation
    - Other EURISCO descriptors can be used for *in situ* CWR data → slightly wider interpretation
    - Additional status terms for some of the descriptors
  - 28 descriptors
  - Standard description + Excel template available



eurisco
Finding seeds for the future

Descriptors for uploading *in situ* CWR passport data to EURISCO

2022-11-02

1 Introduction

This descriptor list describes the data exchange format for uploading passport data from the National Inventories for *in situ* CWR to EURISCO.

The descriptors in this list are a selection from those of the *ex situ* format for upload, with the addition of a few. In case the descriptor name or description is deviating from the *ex situ* upload format, this is indicated in the description.

A significant departure from the *ex situ* data exchange format is the concept of an *in situ* CWR population being an accession. As a result, the population identifiers becomes the ACCENUMB in EURISCO, and the managing institute code and name the INSTCODE/INSTNAME, respectively.

The mandatory fields are, similarly to the *ex situ* upload format, NICODE, INSTCODE, ACCENUMB and GENUS. The combination of these fields has to be unique.

2 General formatting rules

The general formatting rules that apply to the *ex situ* data also apply to the *in situ* data:

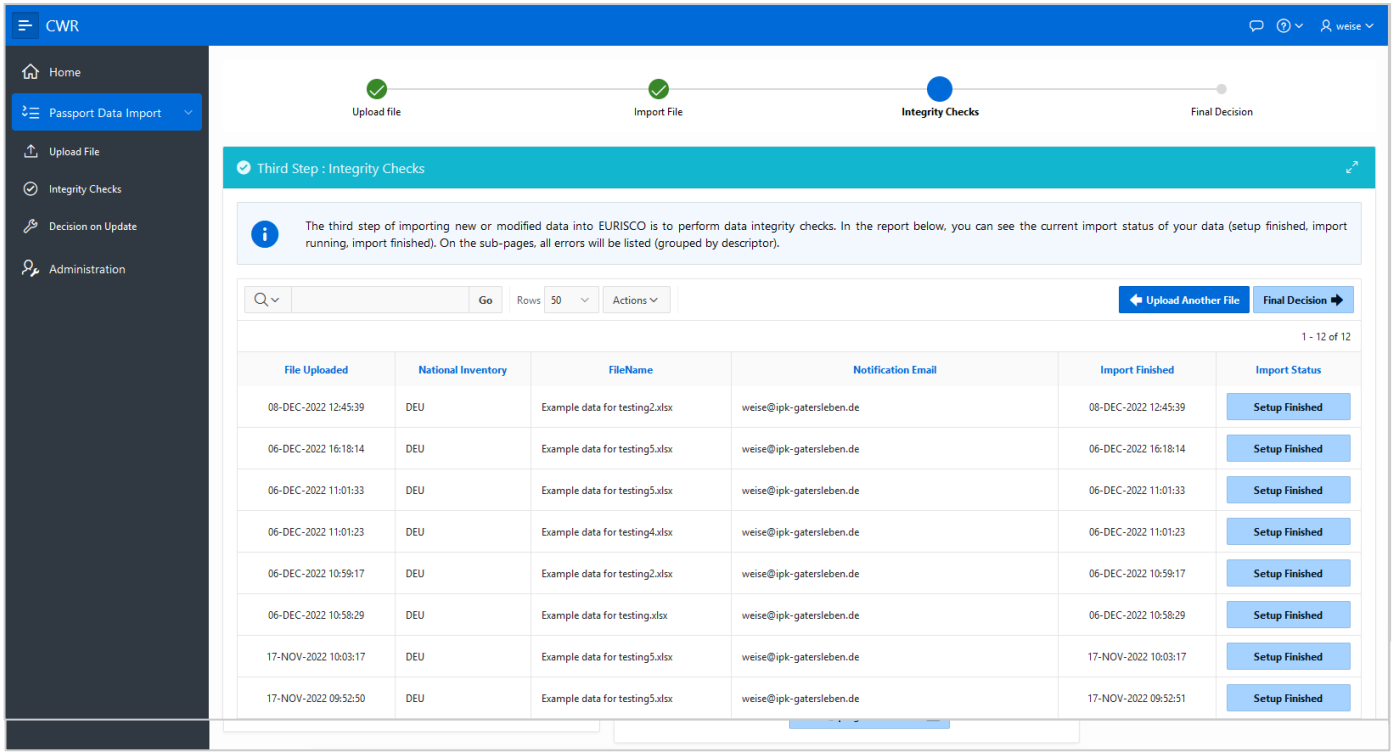
- If a field allows multiple values, these values should be separated by a semicolon (;) without space (e.g. Accession name: Symphony;Emma;Songino).
- A field for which no value is available should be left empty (e.g. Elevation). If data are exchanged in ASCII format, a field with a missing numeric value should be left empty. If data are exchanged in a database format, missing numeric values should be represented by generic NULL values.
- Dates are recorded as YYYYMMDD. If the month or day is missing, this should be indicated with hyphens or '00' [double zero]. If both (month and day) are missing, two double zeros are needed (e.g. 19750000; 197506--, 19750600).
- Country names: Three-letter ISO codes are used for countries. The ISO 3166-1 standard country or area codes are available online at: https://unstats.un.org/unsd/methodology/m49/. **Note**: The list of obsolete codes can be found at: http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3#Reserved_code_elements.
- For institutes, the codes from FAO WIEWS should be used. The current set of institute codes is available from the FAO WIEWS site (http://www.fao.org/wiews).
  - If new institute codes are required, they can be generated online by FAO National Focal Points (http://www.fao.org/agriculture/crops/thematic-sitemap/theme/seeds-pgr/gpa/national-focal- points/en/) or they can be requested from: WIEWS@fao.org.
  - In case no FAO WIEWS code of the institution responsible for, and/or organization that manages the CWR population is available and cannot be generated, the code ('DUMMY') can be used.
  - For institutes that no longer exist, or that were not assigned a FAO WIEWS institute code, please provide full details in the descriptors INSTNAME and LIAISONNAME, respectively.

Recent developments

- Development of upload mechanism
  - Preparation of MS Excel file using the EURISCO *in situ* CWR data standard
  - Upload via the EURISCO intranet
  - Upload by CWR Focal Point
  - Pure web-based upload of data
    - No third-party software needed
    - No specific firewall settings needed

Recent developments

- Interface separated from *ex situ* data upload, but same design
- As simple as possible



Recent developments

- Database schema extension
- Development of procedures for…
  - …data integrity checks
  - …data integration

Still ongoing

- Web interface extension according to user requirements

- Training of country Focal Points in charge of *in situ* CWR data
  - Online workshops/webinars
  - Helpdesk function

- Public awareness products → ECPGR Secretariat

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
NICODE	**National Inventory code** 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
PUID ★	**Persistent unique identifier (PUID)** Persistent, unique identifier (preferably a DOI) assigned to the accession to unambiguously reference it at the global level. NOTE: 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 potentially available under the terms of the MLS).
INSTCODE	**Institute code** FAO WIEWS code of the institution responsible for, and/or organization that manages the CWR population (e.g. protected area authority, nature reserve manager, national park manager, private landowner, etc.).

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
★ **INSTNAME**	**Institute name** Name and short address of the organisation managing the CWR population (e.g. protected area authority, nature reserve manager, national park manager, private landowner, etc.). This descriptor should be used only if INSTCODE has the value ('DUMMY') because the FAO WIEWS code for this institute is not available. NOTE: This descriptor is new and did not occur in the EURISCO format yet.
★ **ACCENUMB**	**Accession number** (≙ population ID) Unique identifier for CWR populations maintained *in situ*. Assigned by the organisation managing the population.
★ **LIAISONCODE**	**Liaison institute code** FAO WIEWS code of the institution that can liaise between the organisation managing the CWR population and the interested user. NOTE: This descriptor is new and did not occur in the EURISCO format yet.
★ **LIAISONNAME**	**Liaison institute name** Name and brief address of the liaison institution in the case that no FAO WIEWS code exists. NOTE: This descriptor is new and did not occur in the EURISCO format yet.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
GENUS	**Genus** Genus name for taxon. Initial uppercase letter required.
SPECIES	**Species** Specific epithet portion of the scientific name in lowercase letters. Only the following abbreviation is allowed: 'sp.'
SPAUTHOR	**Species authority** Provide the authority for the species name.
SUBTAXA ★	**Subtaxon** Subtaxon can be used to store any additional taxonomic identifier. The following abbreviations are allowed: 'subsp.' (for subspecies); 'var.' (for variety); 'f.' (for form). NOTE: The description was slightly modified as cultivar groups cannot occur in CWR populations.
SUBTAUTHOR	**Subtaxon authority** Subtaxon authority at the most detailed taxonomic level.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
★ **ACQDATE**	**Observation date** [YYYYMMDD] The most recent date the population was observed, where YYYY is the year, MM is the month and DD is the day. Missing data (MM or DD) should be indicated with hyphens or '00' [double zero]. NOTE: The name and description of this descriptor have been changed to apply to CWR *in situ*.
★ **ORIGCTY**	**Country of occurrence** Three-letter ISO 3166-1 code of the country where the CWR population was observed or inventoried. NOTE: The name and description of this descriptor have been changed to apply to CWR *in situ*.
★ **COLLSITE**	**Location of occurrence site** Location information below the country level where the population sample was observed. This might include the distance in km and direction from the nearest town, village or map grid reference point (e.g. 7km east of Wageningen in the province of Gelderland). NOTE: The name and description of this descriptor have been changed to apply to CWR *in situ*.
ELEVATION	**Elevation of site** [masl] Elevation of site expressed in metres above sea level. Negative values are allowed.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
★ **DECLATITUDE**	**Latitude of occurrence site** Latitude expressed in decimal degrees. Positive values are north of the Equator; negative values are south of the Equator (e.g. -44.6975). NOTE: The name of this descriptor has been changed to apply to CWR *in situ*. The accuracy of this information that is going to be disseminated may be adjusted as considered appropriate by each country.
★ **DECLONGITUDE**	**Longitude of occurrence site** Longitude expressed in decimal degrees. Positive values are east of the Greenwich Meridian; negative values are west of the Greenwich Meridian (e.g. +120.9123). NOTE: The name of this descriptor has been changed to apply to CWR *in situ*. The accuracy of this information that is going to be disseminated may be adjusted as considered appropriate by each country.
★ **COORDUNCERT**	**Coordinate uncertainty** [m] Uncertainty associated with the coordinates in metres. Leave the value empty if the uncertainty is unknown. Can also be used to indicate the size of the distribution area of the CWR. NOTE: The description of this descriptor has been changed to apply to CWR *in situ*. The coordinate uncertainty should be adjusted if the accuracy of the geographic coordinates is reduced.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
POPSRC	**Status of occurrence site** 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. 10: Wild 11: Forest or woodland, 12: Shrubland, 13: Grassland, 14: Desert or tundra, 15: Aquatic habitat 20: Farm or cultivated area 21: Field, 22: Orchard, 23: Backyard, kitchen or home garden, 24: Fallow land, 25: Pasture, 28: Park 60: Weedy, disturbed or ruderal habitat 61: Roadside, 62: Field margin 99: Other (elaborate in REMARKS field) NOTE: This descriptor is new and did not occur in the EURISCO format yet.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
SITEPROT	**Site protection** 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 (OECM) NOTE: This descriptor is new and did not occur in the EURISCO format yet.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
CONSACTION	**Conservation actions in place** 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: This descriptor is new and did not occur in the EURISCO format yet.

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
SAMPSTAT	**Biological status of accession** The coding scheme proposed can be used at two different levels of detail: either by using the general codes (in boldface) such as 100, 200, or by using the more specific codes such as 110, 120, etc. 100) Wild 110) Natural 120) Semi-natural/wild 130) Semi-natural/sown 200) Weedy 999) Other (Elaborate in REMARKS field) NOTE: The description of this descriptor has changed (less allowed values).

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
OTHERNUMB	**Other identifiers associated with the accession** The identifier(s) of any sample of this population in an *ex situ* collection. Use the following format: INSTCODE:ACCENUMB;INSTCODE:identifier;… INSTCODE and identifier are separated by a colon without space. Pairs of INSTCODE and identifier are separated by a semicolon without space. When the institute is not known, the identifier should be preceded by a colon.
STORAGE ★	Type of germplasm storage For *in situ* CWR populations, this descriptor should always have the value 60. 60) *in situ* wild population NOTE: Status 60 is a new status!
MLSSTAT ★	**MLS status of the accession** The status of the *in situ* accession of the CWR population with regards to the Multilateral System of Access and Benefit-Sharing (MLS) of the International Treaty, if available. 0: Not available under the MLS 1: Available under the MLS

Descriptors for uploading *in situ* CWR passport data

Descriptor	Description
REMARKS	**Remarks** 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. COLLSRC:riverside). Distinct remarks referring to different fields are separated by semicolons without space.
ACCEURL	**Accession URL** URL linking to additional data about the population. Example: http://gbis.ipk-gatersleben.de/gbis_i/detail.jsf?akzessionId=31805 NOTE: This description deviates from the *ex situ* upload format.