

Descriptors for uploading passport data to EURISCO

2017-11-13

1 Introduction

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

The list is an extension of the FAO/Bioversity multi-crop passport descriptors (MCPD), version 2.1, which were published December 2015, developed jointly by Bioversity and FAO, with input from many documentation specialists worldwide, to provide international standards to facilitate germplasm passport information exchange. All MCPDs are included, without change and with the same formatting rules, in the current list.

Four additional descriptors (00, 29-31) were added for the specific purposes of EURISCO.

2 General formatting rules

The following formatting rules, as copied from the MCPD documentation, apply to all fields:

- 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 are 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. 1975----, 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.
 - For institutes that no longer exist, or that were not assigned an FAO WIEWS institute code, please provide full details in descriptors 4.1, 4.1.1, 18.1, 22.1 and 25.1.

3 Descriptors

The descriptors are numbered according to the FAO/Bioversity multi-crop passport descriptors (MCPD), version 2.1. The first descriptor (numbered 00) and the last three (numbered 29-31) are additional, and specific to this EURISCO Descriptor List. Only the four fields identifying the accession are mandatory, all other fields are highly recommended. The mandatory fields are NICODE (00), INSTCODE (1), ACCENUMB (2) and GENUS (5). The combination of these fields has to be unique.

MCPD + EURISCO DESCRIPTORS

00.	National Inventory code	(NICODE)
<p>Code identifying the National Inventory; the code of the country preparing the National Inventory. Exceptions are possible, if agreed with EURISCO, such as NGB.</p> <p>Example: NLD</p>		
0.	Persistent unique identifier	(PUID)
<p>Any persistent, unique identifier assigned to the accession so it can be unambiguously referenced at the global level and the information associated with it harvested through automated means. Report one PUID for each accession.</p> <p style="padding-left: 40px;">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 (http://www.fao.org/plant-treaty/areas-of-work/global-information-system/doi/en/).</p>		
1.	Institute code	(INSTCODE)
<p>FAO WIEWS code of the institute where the accession is maintained. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus a number (e.g. COL001). The current set of institute codes is available from http://www.fao.org/wiews. For those institutes not yet having an FAO Code, or for those with 'obsolete' codes, see '<i>General formatting rules</i>'.</p>		
2.	Accession number	(ACCENUMB)
<p>This is the unique identifier for accessions within a genebank, and is assigned when a sample is entered into the genebank collection (e.g. 'PI 113869').</p>		
3.	Collecting number	(COLLNUMB)
<p>Original identifier assigned by the collector(s) of the sample, normally composed of the name or initials of the collector(s) followed by a number (e.g. 'FM9909'). This identifier is essential for identifying duplicates held in different collections.</p>		
4.	Collecting institute code	(COLLCODE)
<p>FAO WIEWS code of the institute collecting the sample. If the holding institute has collected the material, the collecting institute code (COLLCODE) should be the same as the holding institute code (INSTCODE). Follows INSTCODE standard. Multiple values are separated by a semicolon without space.</p>		
4.1	Collecting institute name	(COLLNAME)
<p>Name of the institute collecting the sample. This descriptor should be used only if COLLCODE cannot be filled because the FAO WIEWS code for this institute is not available. Multiple values are separated by a semicolon without space.</p>		
4.1.1	Collecting institute address	(COLLINSTADDRESS)
<p>Address of the institute collecting the sample. This descriptor should be used only if COLLCODE cannot be filled since the FAO WIEWS code for this institute is not available. Multiple values are separated by a semicolon without space.</p>		

4.2 Collecting mission identifier	(COLLMISSID)
Identifier of the collecting mission used by the Collecting Institute (4 or 4.1) (e.g. 'CIATFOR-052', 'CN426').	
5. Genus	(GENUS)
Genus name for taxon. Initial uppercase letter required.	
6. Species	(SPECIES)
Specific epithet portion of the scientific name in lowercase letters. Only the following abbreviation is allowed: 'sp.'	
7. Species authority	(SPAUTHOR)
Provide the authority for the species name.	
8. Subtaxon	(SUBTAXA)
Subtaxon can be used to store any additional taxonomic identifier. The following abbreviations are allowed: 'subsp.' (for subspecies); 'convar.' (for convariety); 'var.' (for variety); 'f.' (for form); 'Group' (for 'cultivar group').	
9. Subtaxon authority	(SUBTAUTHOR)
Provide the subtaxon authority at the most detailed taxonomic level.	
10. Common crop name	(CROPNAME)
Common name of the crop. Example: 'malting barley', 'macadamia', 'maïs'.	
11. Accession name	(ACCENAME)
Either a registered or other designation given to the material received, other than the donor's accession number (23) or collecting number (3). First letter uppercase. Multiple names are separated by a semicolon without space. Example: Accession name: Bogatyr;Symphony;Emma.	
12. Acquisition date [YYYYMMDD]	(ACQDATE)
Date on which the accession entered the collection 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].	
13. Country of origin	(ORIGCTY)
3-letter ISO 3166-1 code of the country in which the sample was originally collected.	

Note: Descriptors 14 to 16 below should be completed accordingly only if it was 'collected'.

14. Location of collecting site	(COLLSITE)
Location information below the country level that describes where the accession was collected, preferable in English. This might include the distance in kilometres and direction from the nearest town, village or map grid reference point, (e.g. 7 km south of Curitiba in the state of Parana).	
15. Geographical coordinates	
<ul style="list-style-type: none"> • For latitude and longitude descriptors, two alternative formats are proposed, but the one reported by the collecting mission should be used • Latitude and longitude in decimal degree format with a precision of four decimal places corresponds to approximately 10 m at the Equator and describes the point-radius representation of the location, along with Geodetic datum and Coordinate uncertainty in metres. 	

Note: The following two mutually exclusive formats can be used for latitude:

15.1 Latitude of collecting site (Decimal degrees format)	(DECLATITUDE)
Latitude expressed in decimal degrees. Positive values are North of the Equator; negative values are South of the Equator (e.g. -44.6975).	
15.2 Latitude of collecting site (Degrees, Minutes, Seconds format)	(LATITUDE)
Degrees (2 digits) minutes (2 digits), and seconds (2 digits) followed by N (North) or S (South) (e.g. 103020S). Every missing digit (minutes or seconds) should be indicated with a hyphen. Leading zeros are required (e.g. 10----S; 011530N; 4531--S).	

Note: The following two mutually exclusive formats can be used for longitude:

15.3 Longitude of collecting site (Decimal degrees format)	(DECLONGITUDE)
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).	
15.4 Longitude of collecting site (Degrees, Minutes, Seconds format)	(LONGITUDE)
Degrees (3 digits), minutes (2 digits), and seconds (2 digits) followed by E (East) or W (West) (e.g. 0762510W). Every missing digit (minutes or seconds) should be indicated with a hyphen. Leading zeros are required (e.g. 076----W).	
15.5 Coordinate uncertainty [m]	(COORDUNCERT)
Uncertainty associated with the coordinates in metres. Leave the value empty if the uncertainty is unknown.	
15.6 Coordinate datum	(COORDDATUM)
The geodetic datum or spatial reference system upon which the coordinates given in decimal latitude and decimal longitude are based (e.g. WGS84, ETRS89, NAD83). The GPS uses the WGS84 datum.	
15.7 Georeferencing method	(GEOREFMETH)
The georeferencing method used (GPS, determined from map, gazetteer, or estimated using software). Leave the value empty if georeferencing method is not known.	
16. Elevation of collecting site [masl]	(ELEVATION)
Elevation of collecting site expressed in metres above sea level. Negative values are allowed.	
17. Collecting date of sample [YYYYMMDD]	(COLLDATE)
Collecting date of the sample, 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].	
18. Breeding institute code	(BREDCODE)
FAO WIEWS code of the institute that has bred the material. If the holding institute has bred the material, the breeding institute code (BREDCODE) should be the same as the holding institute code (INSTCODE). Follows INSTCODE standard. Multiple values are separated by a semicolon without space.	
18.1 Breeding institute name	(BREDNAME)
Name of the institute (or person) that bred the material. This descriptor should be used only if BREDCODE cannot be filled because the FAO WIEWS code for this institute is not available. Multiple names are separated by a semicolon without space.	

19. Biological status of accession**(SAMPSTAT)**

The coding scheme proposed can be used at 3 different levels of detail: either by using the general codes (in boldface) such as 100, 200, 300, 400, 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**300) Traditional cultivar/landrace****400) Breeding/research material**

410) Breeder's line

411) Synthetic population

412) Hybrid

413) Founder stock/base population

414) Inbred line (parent of hybrid cultivar)

415) Segregating population

416) Clonal selection

420) Genetic stock

421) Mutant (e.g. induced/insertion mutants, tilling populations)

422) Cytogenetic stocks (e.g. chromosome addition/substitution, aneuploids, amphiploids)

423) Other genetic stocks (e.g. mapping populations)

500) Advanced or improved cultivar (conventional breeding methods)**600) GMO** (by genetic engineering)**999) Other** (Elaborate in REMARKS field)**20. Ancestral data****(ANCEST)**

Information about either pedigree or other description of ancestral information (e.g. parent variety in case of mutant or selection). For example a pedigree 'Hanna/7*Atlas//Turk/8*Atlas' or a description 'mutation found in Hanna', 'selection from Irene' or 'cross involving amongst others Hanna and Irene'.

<p>21. Collecting/acquisition source</p> <p>The coding scheme proposed can be used at 2 different levels of detail: either by using the general codes (in boldface) such as 10, 20, 30, 40, etc., or by using the more specific codes, such as 11, 12, etc.</p> <p>10) Wild habitat</p> <ul style="list-style-type: none"> 11) Forest or woodland 12) Shrubland 13) Grassland 14) Desert or tundra 15) Aquatic habitat <p>20) Farm or cultivated habitat</p> <ul style="list-style-type: none"> 21) Field 22) Orchard 23) Backyard, kitchen or home garden (urban, peri-urban or rural) 24) Fallow land 25) Pasture 26) Farm store 27) Threshing floor 28) Park <p>30) Market or shop</p> <p>40) Institute, Experimental station, Research organization, Genebank</p> <p>50) Seed company</p> <p>60) Weedy, disturbed or ruderal habitat</p> <ul style="list-style-type: none"> 61) Roadside 62) Field margin <p>99) Other (Elaborate in REMARKS field)</p>	<p>(COLLSRC)</p>
<p>22. Donor institute code</p> <p>FAO WIEWS code of the donor institute. Follows INSTCODE standard.</p>	<p>(DONORCODE)</p>
<p>22.1 Donor institute name</p> <p>Name of the donor institute (or person). This descriptor should be used only if DONORCODE cannot be filled because the FAO WIEWS code for this institute is not available.</p>	<p>(DONORNAME)</p>
<p>23. Donor accession number</p> <p>Identifier assigned to an accession by the donor. Follows ACCENUMB standard.</p>	<p>(DONORNUMB)</p>
<p>24. Other identifiers associated with the accession</p> <p>Any other identifiers known to exist in other collections for this accession. 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.</p>	<p>(OTHERNUMB)</p>
<p>25. Location of safety duplicates</p> <p>FAO WIEWS code of the institute(s) where a safety duplicate of the accession is maintained. Multiple values are separated by a semicolon without space. Follows INSTCODE standard.</p>	<p>(DUPLSITE)</p>
<p>25.1 Institute maintaining safety duplicates</p> <p>Name of the institute where a safety duplicate of the accession is maintained. Multiple values are separated by a semicolon without space.</p>	<p>(DUPLINSTNAME)</p>

<p>26. Type of germplasm storage</p> <p>If germplasm is maintained under different types of storage, multiple choices are allowed, separated by a semicolon (e.g. 20;30). (Refer to FAO/IPGRI Genebank Standards 1994 for details on storage type.)</p> <p>10) Seed collection</p> <p>11) Short term 12) Medium term 13) Long term</p> <p>20) Field collection</p> <p>30) <i>In vitro</i> collection</p> <p>40) Cryopreserved collection</p> <p>50) DNA collection</p> <p>99) Other (elaborate in REMARKS field)</p>	<p>(STORAGE)</p>
<p>27. MLS status of the accession</p> <p>The status of the PGRFA with regard to the Multilateral System of Access and Benefit-Sharing (MLS) of the International Treaty on Plant Genetic Resources for Food and Agriculture. Leave the value empty if the status is not known.</p> <p>0 No (not available under the MLS) 1 Yes (available under the MLS)</p>	<p>(MLSSTAT)</p>
<p>28. Remarks</p> <p>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.</p>	<p>(REMARKS)</p>
<p>29. Accession URL</p> <p>URL linking to additional data about the accession either in the holding genebank or from another source. Example: http://gbis.ipk-gatersleben.de/gbis_i/detail.jsf?akzessionId=31805</p>	<p>(ACCEURL)</p>
<p>30. AEGIS status</p> <p>The coded status of an accession with regards to <i>A European Genebank Integrated System</i> (AEGIS). Provides the information, whether the accession is conserved for AEGIS.</p> <p>0 No (not part of the AEGIS) 1 Yes (part of the AEGIS)</p> <p>If the AEGIS status is unknown, the field stays empty.</p>	<p>(AEGISSTAT)</p>
<p>31. Historic accession</p> <p>This flag indicates if an accession is no longer actively maintained by a holding genebank. In this case, the historic data might nonetheless provide valuable information.</p> <p>0 No (actively maintained) 1 Yes (not maintained anymore by the holding genebank)</p> <p>Leave the value empty if the status is not known.</p>	<p>(HISTORIC)</p>

4 List of changes

In the following, the major changes compared with the 2012 version of the EURISCO data exchange format are listed.

4.1 COMMON FORMATTING RULES

- **Missing data:** For those descriptors related to 'Dates', missing data can be indicated with hyphens or '00'[double zero] to ensure consistent data transfer whatever system is used.
- **Latitude and longitude:** Two formats (Latitude: DDMMSSH or +/-DD.DDDD; Longitude: DDDMMSSSE/W or +/-DDD.DDDD) are allowed, but only one of the two coordinate formats should be used (the one that was reported by the collecting mission).

4.2 NEW DESCRIPTORS

- PUID
- COLLINSTADDRESS
- COLLMISSID
- DECLATITUDE
- DECLONGITUDE
- COORDUNCERT
- COORDDATUM
- GEOREFMETH
- HISTORIC

4.3 MODIFIED DESCRIPTORS

- **COLLCODE:** multiple values allowed
- **DUPLSITE:** multiple values allowed
- **BREDCODE:** multiple values allowed
- **COLLNAME:** replaces COLLDESCR
- **BREDNAME:** replaces BREDESCR
- **DONORNAME:** replaces DONORDESCR
- **DUPLINSTNAME:** replaces DUPLDESCR