

Data exchange standard for uploading characterisation and evaluation data from National Inventories to EURISCO

This document describes the data collection templates for submitting characterisation & evaluation (C&E) data to the European search catalogue for plant genetic resources (EURISCO). The templates are based on the EPGRIS3 document “Inclusion of C&E data in EURISCO – analysis and options”¹ by Theo van Hintum (May 2009).

Date: 2015-08-07

Version: 1.0

Content

1	Data model	2
2	Templates for C&E data collection	2
2.1	DATASET	2
2.2	EXPERIMENT	3
2.3	TRAIT	3
2.4	GENOTYPE	4
2.5	SCORE	4

¹ [Inclusion of C&E data in EURISCO – analysis and options](#)

1 Data model

The data model for the EURISCO C&E data consists of five components: GENOTYPE, SCORE, TRAIT, EXPERIMENT and DATASET. A DATASET may comprise different EXPERIMENTs. Within an EXPERIMENT, multiple GENOTYPEs are SCOREd for different TRAITs. A GENOTYPE is identified by the unique combination of the four EURISCO descriptors NICODE, INSTCODE, ACCENUMB and GENUS. The connections between the components are illustrated by Fig. 1.

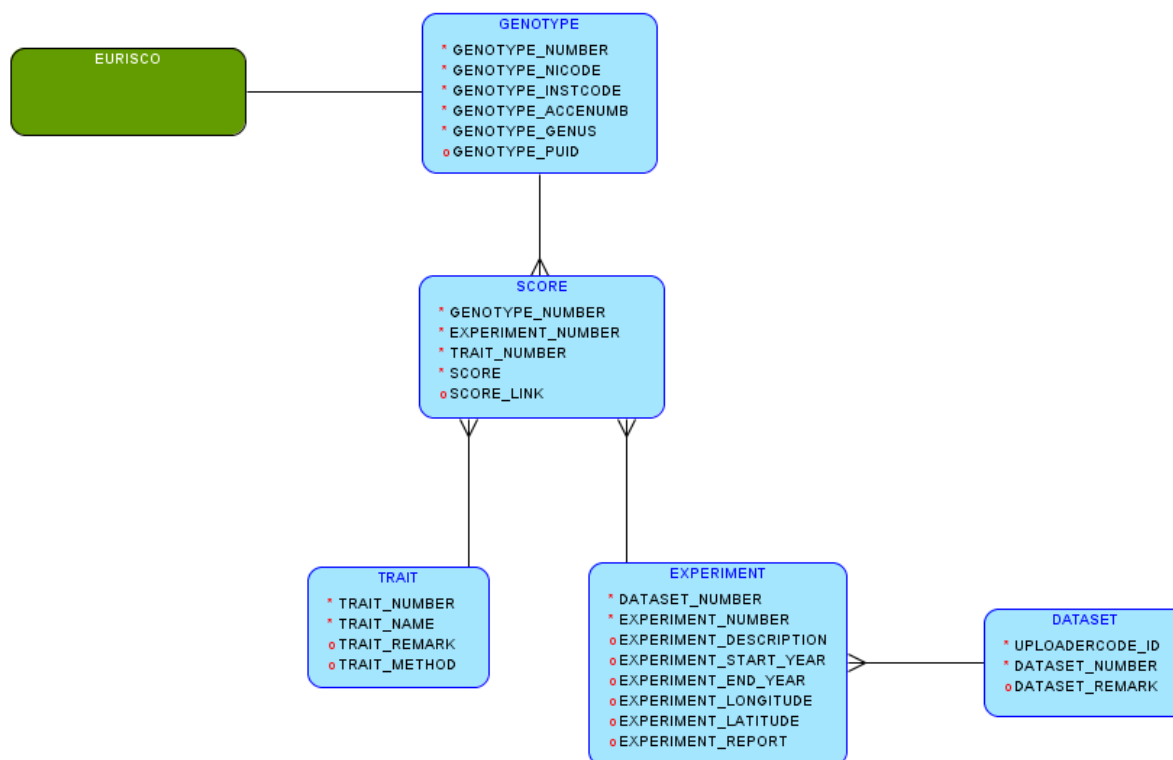


Fig. 1: Simplified data model for C&E data in EURISCO. Mandatory fields are marked by an asterisk (*).

2 Templates for C&E data collection

In the following, the five templates for DATASET, EXPERIMENT, TRAIT, GENOTYPE and SCORE are described and illustrated with an example each. The templates will be made available as MS-Excel files.

Only non-confidential C&E data of accessions listed in EURISCO may be uploaded.

At least, experiment descriptions, trait names and method descriptions should be in English language. However, it is recommended to use the English language for all attributes.

2.1 DATASET

In order to upload data about multiple experiments at once, the experiments should be packaged by a dataset. The dataset template comprises the following fields:

- **UPLOADERCODE:** ID code, provided by EURISCO, for the registered authorised data provider uploading the data to EURISCO (*mandatory*).
- **DATASET_NUMBER:** Unique number necessary to link the EXPERIMENTs with the DATASET. This number should be unique and persistent for the data provider (*mandatory*).
- **DATASET_REMARK:** Any general remark relevant to all scores in the dataset (*max. 2000 alphanumeric characters*).

UPLOADERCODE	DATASET_NUMBER	DATASET_REMARK
DEU	3	This dataset contains C&E data of Allium species.
...

2.2 EXPERIMENT

The experiment template comprises all meta information, which can help to interpret the C&E data, such as experiment set-up, weather or soil conditions, experiment location etc. The experiment template comprises the following fields:

- **DATASET_NUMBER:** Key to DATASET (*mandatory*).
- **EXPERIMENT_NUMBER:** Unique number in the dataset for the experiment; this number should be unique and persistent for the data provider (*mandatory*).
- **EXPERIMENT_DESCRIPTION:** Brief English description of the experiment. Information relevant for the interpretation of the scores in the experiment, such as experimental design, experimenter, weather, etc. (*max. 2000 alphanumeric characters*).
- **EXPERIMENT_START_YEAR:** The year the experiment was performed (started) (*4 numeric characters*).
- **EXPERIMENT_END_YEAR:** The year in which the experiment ended (*4 numeric characters*).
- **EXPERIMENT_LONGITUDE:** The longitude of the experimental site, provided it was an experiment in the open field (*decimal number*).
- **EXPERIMENT_LATITUDE:** The latitude of the experimental site, provided it was an experiment in the open field (*decimal number*).
- **EXPERIMENT_REPORT:** A reference to the report of the experiment, either supplied with the data, or the URL of the report or original data. It will be possible to upload an additional file containing the report. In the latter case, only the file name should be given here (*max. 255 alphanumeric characters*).

DATASET_NUMBER	EXPERIMENT_NUMBER	EXPERIMENT_DESCRIPTION	EXPERIMENT_START_YEAR	EXPERIMENT_END_YEAR	EXPERIMENT_LONGITUDE	EXPERIMENT_LATITUDE	EXPERIMENT_REPORT
3	12	Elicitation of characterisation data for Allium sativum	1998		11.278414	51.826059	http://...
3	13	Elicitation of characterisation data for Allium sativum	2004	2005	11.278414	51.826059	http://...
...

2.3 TRAIT

The trait template should be used to describe the phenotypic traits as well as the methods, which were used for scoring. The trait template comprises the following fields:

- **TRAIT_NUMBER:** Unique, temporary number of the trait in the dataset (*mandatory*).
- **TRAIT_NAME:** English name of the trait (*max. 100 alphanumeric characters, mandatory*).
- **TRAIT_REMARK:** Any general remark that helps to interpret the trait (*max. 2000 alphanumeric characters*).

- **TRAIT_METHOD:** An English description of the method for measuring and the scale used (*max. 2000 alphanumeric characters*).

TRAIT_NUMBER	TRAIT_NAME	TRAIT_REMARK	TRAIT_METHOD
1	Ability to produce scape	Modified after IPGRI 7.2.2.	Rating score from 0 (absent) to 5 (long scape, bulbils top only)
2	Foliage colour	Modified after IPGRI 7.1.1.	Rating score (light green (3), medium green (5), dark green (7))
...

2.4 GENOTYPE

The genotype template should list all accessions for which C&E data will be uploaded. As a prerequisite, these accessions have to be already listed in EURISCO. The genotype template comprises the following fields:

- **GENOTYPE_NUMBER:** Unique temporary number for the genotype within the dataset (*mandatory*).
- **GENOTYPE_NICODE:** Part of the unique key of EURISCO (*mandatory*).
- **GENOTYPE_INSTCODE:** Part of the unique key of EURISCO (*mandatory*).
- **GENOTYPE_ACCENUMB:** Part of the unique key of EURISCO (*mandatory*).
- **GENOTYPE_GENUS:** Part of the unique key of EURISCO (*mandatory*).
- **GENOTYPE_PUID:** Permanent unique identifier (PUID) of a genebank accession (*placeholder*). *PUIDs are necessary for unambiguously and permanently identifying genebank accessions. Since the discussions about this issue are still ongoing in the PGR community, this field serves only as a placeholder. For more information see <http://www.planttreaty.org/doi>.*

GENOTYPE_NUMBER	GENOTYPE_NICODE	GENOTYPE_INSTCODE	GENOTYPE_ACCENUMB	GENOTYPE_GENUS	PUID
1	DEU	DEU146	ALL 100	Allium	
2	DEU	DEU146	ALL 1003	Allium	
...	

2.5 SCORE

Finally, the phenotypic values of the certain accessions should be compiled in the score template. The score template comprises the following fields:

- **GENOTYPE_NUMBER:** Key to GENOTYPE (*mandatory*).
- **EXPERIMENT_NUMBER:** Key to EXPERIMENT (*mandatory*).
- **TRAIT_NUMBER:** Key to TRAIT (*mandatory*).
- **SCORE:** Actual score (*max. 10 alphanumeric characters, mandatory*).
- **SCORE_LINK:** Link to a publication on accession level (*max. 255 alphanumeric characters*).

GENOTYPE_NUMBER	EXPERIMENT_NUMBER	TRAIT_NUMBER	SCORE	SCORE_LINK
1	12	1	5	http://...
1	12	2	7	http://...
1	13	1	5	http://...
1	13	2	5	http://...
...