

ESTABLISHMENT OF AN EUROPEAN PLANT GENETIC RESOURCES INFORMATION INFRA-STRUCTURE



Draft 2002-06-28

EURISCO data check and validation. Presentation for 2nd regional meetings (2002).

Follows data definitions for Multi Crop Passport Descriptors.

TODO:

- * Elaborate with examples on the format of the feedback noted as 'notify' in table 1.
- * Discuss action for the errors noted as 'discard record' in table 1. Can we discard records?
- * Can we produce an online data validation with immediate feedback

Table 1. Suggested list of EURISCO data validation checks to be performed.

tno	dno	descriptor	principle	check	action
1	0	global	unique	combination of nicode, instcode, accenumb, genus need to be unique	discard record, notify
2	0	global	mandatory	empty fields nicode, instcode, accenumb, genus not allowed	discard record, notify
3	12	ACQDATE	date	fields of datatype date need valid date format (YYYYMMDD) with unknowns dashed	notify
4	18	COLLDATE	date	fields of datatype date need valid date format (YYYYMMDD) with unknowns dashed	notify
5	1	INSTCODE	instcode	only valid WIEWS institute codes accepted	discard record, notify
6	4	COLLCODE	instcode	only valid WIEWS institute codes accepted	notify
7	19	BREDCODE	instcode	only valid WIEWS institute codes accepted	notify
8	23	DONORCODE	instcode	only valid WIEWS institute codes accepted	notify
9	5	GENUS	taxon	check valid taxon, synonyms accepted (vs GRIN ?)	notify
10	6	SPECIES	taxon	check valid taxon, synonyms accepted (vs GRIN ?)	notify
11	7	SPAUTHOR	taxon	check valid taxon, synonyms accepted (vs GRIN ?)	notify
12	8	SUBTAXA	taxon	check valid taxon, synonyms accepted (vs GRIN ?)	notify
13	9	SUBTAUTHOR	taxon	check valid taxon, synonyms accepted (vs GRIN ?)	notify
14	13	ORIGCTY	isocou	require valid ISO 3166-1 code, old codes accepted	notify
15	15	LATITUDE	coordinate	require valid degree (0-90), minute (0-59), second (0-59), hemisphere (N,S)	discard field, notify
16	16	LONGITUDE	coordinate	require valid degree (0-180), minute (0-59), second (0-	discard field,

				59), hemisphere (W,E)	notify
17	17	ELEVATION	integer	require data type integer	discard field, notify
18	20	SAMPSTAT	mcpcode	valid MCPDL SAMPSTAT code	notify
19	22	COLLSRC	mcpcode	valid MCPDL COLLSRC code	notify
20	27	STORAGE	mcpcode	valid MCPDL STORAGE code	notify
21	a1	NI code	nicode	valid NI code	discard record/dataset? , notify
22	a2	accession link	url	valid URL	notify
23	a3	record date	update date	update date	if not valid use 'today'

tno – test number, **dno** - mcp descriptor number, **principle** – short name of principle (could use same sub routine), **check** – requirements to be checked, **action** –action to be taken

Table 2. EURISCO descriptors listed together with a note on which validation check to be performed.

MCPDL_DESNAM	MCPDL	Check principles
1 Institute code	INSTCODE	mandatory, unique, instcode
2 Accession number	ACCENUMB	mandatory, unique
3 Collecting number	COLLNUMB	
4 Collecting institute code	COLLCODE	instcode
5 Genus	GENUS	taxon
6 Species	SPECIES	taxon
7 Species authority	SPAUTHOR	taxon
8 Subtaxa	SUBTAXA	taxon
9 Subtaxa authority	SUBTAUTHOR	taxon
10 Common crop name	CROPNAME	
11 Accession name	ACCENAME	
12 Acquisition date	ACQDATE	date
13 Country of origin	ORIGCTY	isocou
14 Location of collecting site	COLLSITE	
15 Latitude of collecting site	LATITUDE	coordinate
16 Longitude of collecting site	LONGITUDE	coordinate
17 Elevation	ELEVATION	integer
18 Collecting date of sample	COLLDATE	date
19 Breeding institute code	BREDCODE	instcode
20 Biological status of original sample	SAMPSTAT	mcpcode
21 Ancestral data	ANCEST	
22 Collecting source	COLLSRC	mcpcode
23 Donor code	DONORCODE	instcode
24 Donor number	DONORNUMB	
25 Other identification numbers associated with the accession	OTHERNUMB	
26 Location of safety duplicates	DUPLSITE	instcode
27 Storage duration	STORAGE	mcpcode
28 Remarks	REMARKS	
a1 NI code	nicode	mandatory, unique

a2 accession link
a3 record date

acceurl
recdate

url
date