

Stephan Weise



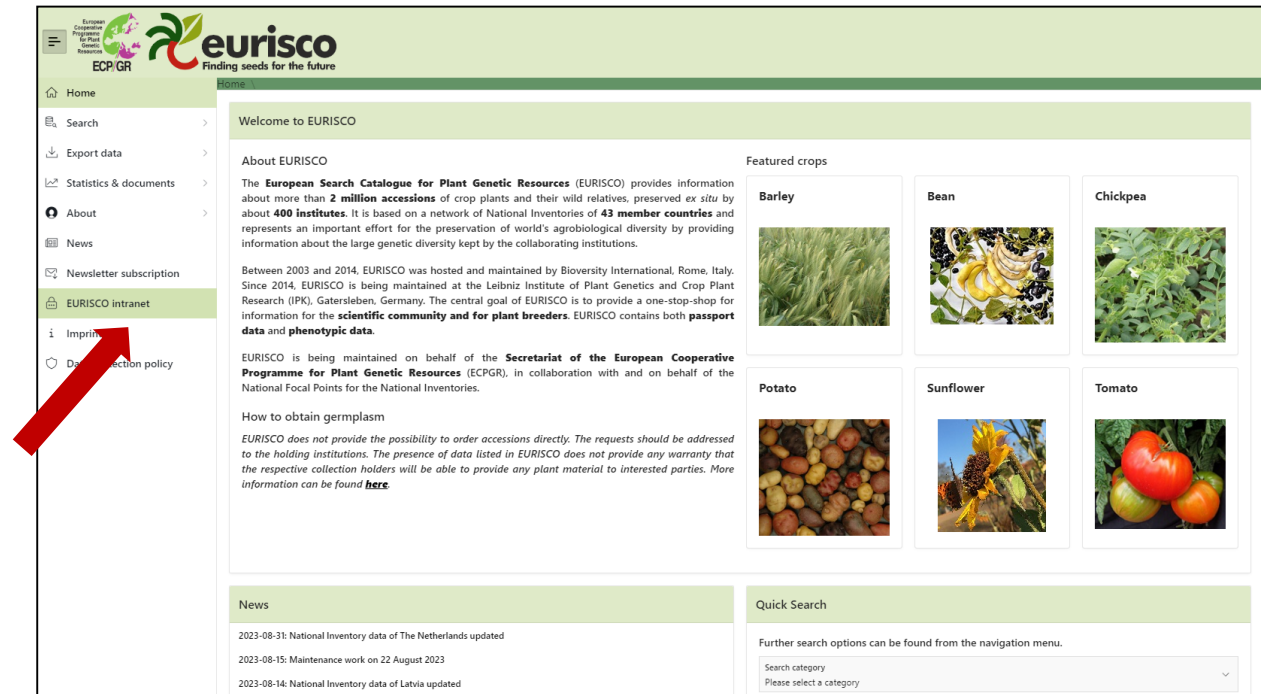
Passport data updates

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



EURISCO intranet

- Import component for National Inventory Focal Points
 - Web interface with Oracle APEX
 - PL/SQL packages for uploading, checking and updating data
 - Implementation of incremental updates



The screenshot shows the EURISCO website interface. The header includes the ECP GR logo and the text "eurisco Finding seeds for the future". The main content area is titled "Welcome to EURISCO" and contains an "About EURISCO" section. A red arrow points to the "EURISCO intranet" link in the left navigation menu. The "Featured crops" section displays images of Barley, Bean, Chickpea, Potato, Sunflower, and Tomato. The "News" section lists recent updates, and the "Quick Search" section provides a search bar and a dropdown menu for search categories.

EURISCO intranet



Step 1
Upload
file

Step 2
Import
File

Step 3
Integrity
check
results

Step 4
Decision
about
update

eurisco

Descriptors for uploading information from National Inventories to EURISCO

Introduction
This document set is used to create a list of data to be uploaded to the EURISCO database. The list is an extension of the Descriptor 2000, developed by the International Working Group on Plant Genetic Resources (IWRG), which is used for the identification and description of plant genetic resources. The list is used to identify the descriptors of a list of descriptors that have been used in the past.

General format rules
Following General rules, as well as:
• If a field allows multiple values, the values should be separated by a semicolon.
• A field is marked as 'Mandatory' if it is required for the upload of a descriptor.
• Dates are recorded as YYYY-MM-DD.
• For writing coordinates, the format is: longitude, latitude, altitude (e.g. 10.000000, 45.000000, 1000.000000).
• For writing numbers, the format is: integer, decimal, scientific notation (e.g. 10, 10.000000, 1e+02).
• The preferred language is English.

Descriptors
The descriptors are numbered according to the Descriptor 2000. Only the field descriptions are provided. The field descriptions are available in the EURISCO database.

1. National inventory code
Code identifying the National Inventory, the code of the country preparing the National Inventory. Example: DE
Example: DE

2. Accession number
The number assigned to a particular accession within the national collection.
Example: 123456789

3. Collecting number
Organization number for the collection.
Example: 123456789

4. Collecting institute code
Code of the institute collecting the sample.
Example: 123456789

5. Name
Name of the accession.
Example: 123456789

6. Species
Species name of the accession.
Example: 123456789

7. Species authority
The authority for the species name.
Example: 123456789

8. Substrate
Substrate used to store any genetic material.
Example: 123456789

9. Accession authority
The authority for the accession.
Example: 123456789

10. Common crop name
Name of the most widespread crop.
Example: 123456789

11. Accession name
Name of the accession.
Example: 123456789

12. Accession date
Date when the accession entered the National Inventory.
Example: 123456789

13. Origin
Origin of the accession.
Example: 123456789

14. Latitude of collecting site
Latitude of the collecting site.
Example: 123456789

15. Longitude of collecting site
Longitude of the collecting site.
Example: 123456789

16. Language of collecting site
Language of the collecting site.
Example: 123456789

17. Evaluation of collecting site
Evaluation of the collecting site.
Example: 123456789

18. Biological value of accession
Biological value of the accession.
Example: 123456789

19. Breeding method code
Code of the breeding method.
Example: 123456789

20. Breeding method name
Name of the breeding method.
Example: 123456789

21. Ancestral data
Information about the program or other description of ancestral information.
Example: 123456789

22. Collecting date
Date when the accession was collected.
Example: 123456789

23. Collecting site name
Name of the collecting site.
Example: 123456789

24. Collecting site code
Code of the collecting site.
Example: 123456789

25. Breeding method name
Name of the breeding method.
Example: 123456789

26. Breeding method code
Code of the breeding method.
Example: 123456789

27. Other identification numbers
Any other identification numbers.
Example: 123456789

28. Accession number
The number assigned to a particular accession within the national collection.
Example: 123456789

29. Accession URL
URL linking additional data about the accession.
Example: 123456789

30. ML3 Status
The status of the accession in the ML3.
Example: 123456789

31. Accession status
The status of the accession in the EURISCO.
Example: 123456789

32. Decoded collecting institute
Decoded name of the collecting institute.
Example: 123456789

33. Decoded breeding institute
Decoded name of the breeding institute.
Example: 123456789

34. Decoded donor institute
Decoded name of the donor institute.
Example: 123456789

35. Decoded safety duplication location
Decoded name of the safety duplication location.
Example: 123456789

36. Accession URL
URL linking additional data about the accession.
Example: 123456789

37. Type of germplasm storage
Type of germplasm storage.
Example: 123456789

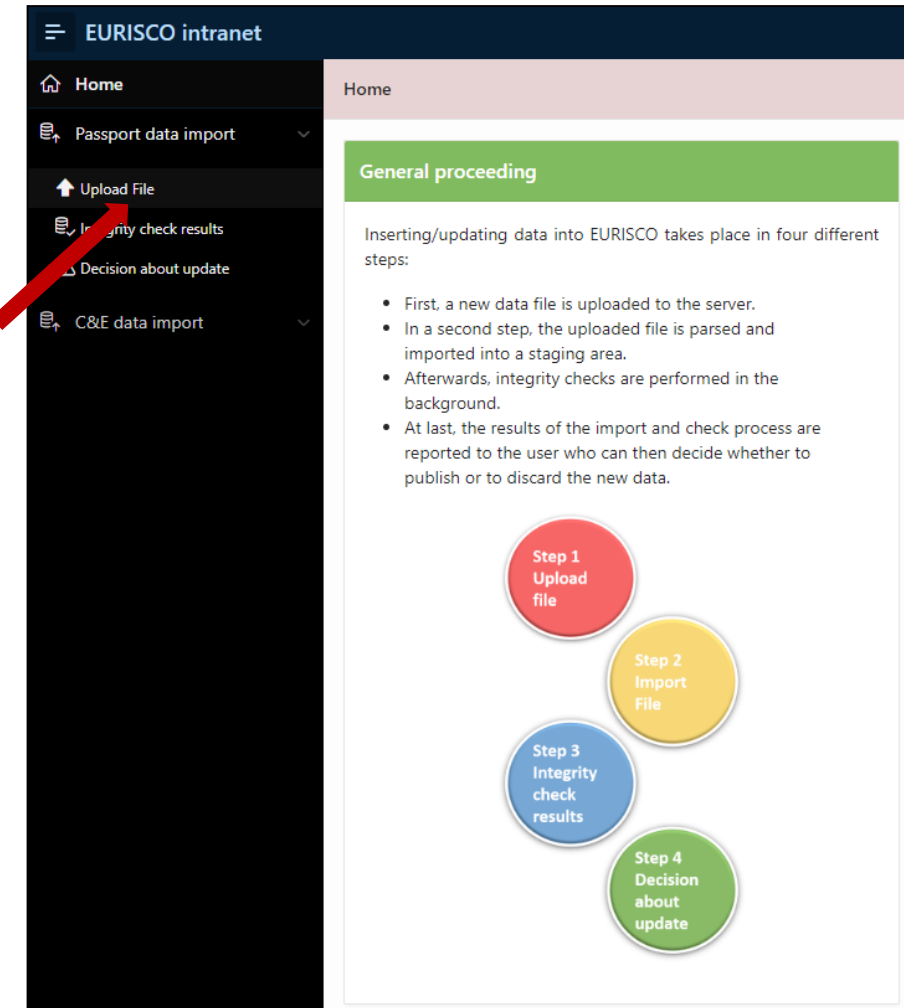
38. ML3 Status
The status of the accession in the ML3.
Example: 123456789

39. Accession status
The status of the accession in the EURISCO.
Example: 123456789

40. Accession status
The status of the accession in the EURISCO.
Example: 123456789

File upload


- Prepare the passport data using the current EURISCO-MCPD format
- Use a MS Excel (.xlsx) file, descriptor names as column headings
- Login
- Expand navigation menu
- Open “Passport data import”
- Click on “Upload file”



The screenshot displays the EURISCO intranet interface. On the left, a dark navigation menu is expanded, showing options: Home, Passport data import, Upload File (highlighted with a red arrow), Integrity check results, Decision about update, and C&E data import. The main content area, titled 'Home', features a green header for 'General proceeding'. Below this, text explains that data insertion/update occurs in four steps, followed by a bulleted list: 1. Upload a new data file to the server; 2. Parse and import the file into a staging area; 3. Perform background integrity checks; 4. Report results for user decision. A circular diagram at the bottom right illustrates these four steps: Step 1 (Upload file), Step 2 (Import File), Step 3 (Integrity check results), and Step 4 (Decision about update).

File upload

- Select file to be uploaded
- Select validation profile
- Enter notification email address
- Start the upload



Home \ Upload File

Final Decision

File must be formatted in accordance with the MCPD

The file

Validation Profile [X]

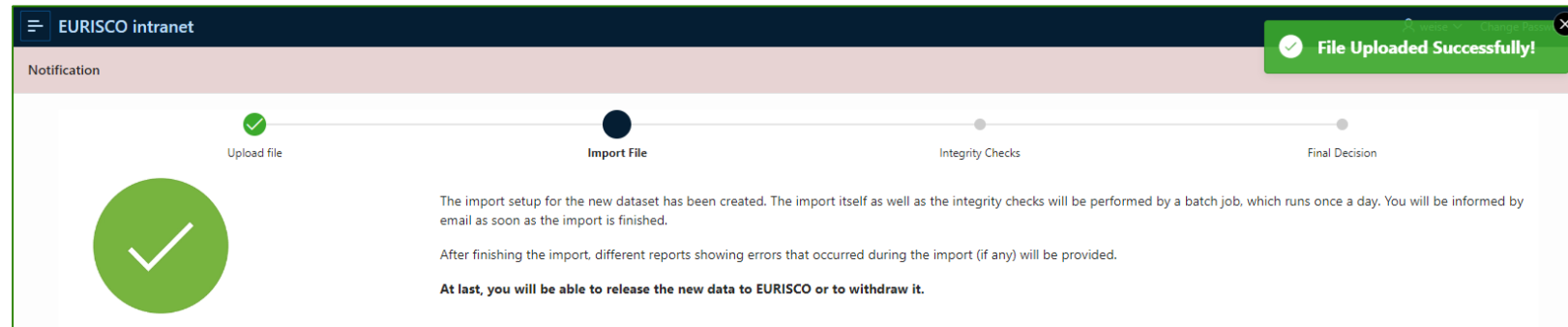
Errors abort session:
If an error occurs, the whole import operation will be aborted.

Records with errors are skipped:
If an error occurs within a record, this record will be skipped. The remaining records will be imported.

Descriptors with errors are skipped:
If a descriptor value within a record contains an error, only this value will be skipped. The remaining descriptor values of the respective record will be imported. However, in case of an error within one of the mandatory descriptors, the whole record will be skipped.

File import

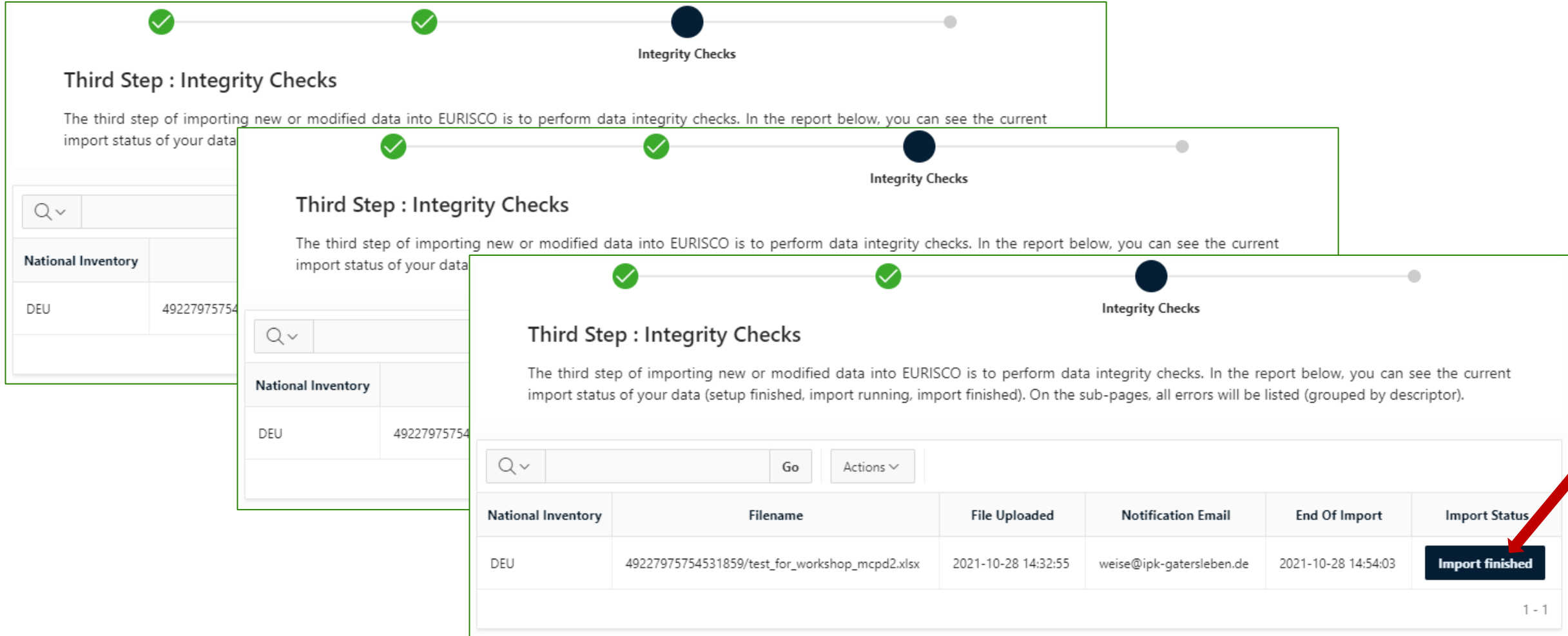
- In case of basic format errors (e.g. non-supported column heading)
→ immediate response
- After successful upload
→ background process
- If file read and data written into staging area
→ start of integrity checks



The screenshot shows the EURISCO intranet interface. At the top right, a green notification box says "File Uploaded Successfully!". Below it, a progress bar indicates the status of the import process. The progress bar has four steps: "Upload file" (completed with a green checkmark), "Import File" (current step, indicated by a black dot), "Integrity Checks" (pending), and "Final Decision" (pending). To the left of the progress bar is a large green circle with a white checkmark. Below the progress bar, there is a notification text: "The import setup for the new dataset has been created. The import itself as well as the integrity checks will be performed by a batch job, which runs once a day. You will be informed by email as soon as the import is finished. After finishing the import, different reports showing errors that occurred during the import (if any) will be provided. At last, you will be able to release the new data to EURISCO or to withdraw it."

No input needed from your side

Data integrity checks



Third Step : Integrity Checks

The third step of importing new or modified data into EURISCO is to perform data integrity checks. In the report below, you can see the current import status of your data

National Inventory	Filename	File Uploaded	Notification Email	End Of Import	Import Status
DEU	49227975754531859/test_for_workshop_mcpd2.xlsx	2021-10-28 14:32:55	weise@ipk-gatersleben.de	2021-10-28 14:54:03	Import finished

1 - 1

Data integrity checks

Q v Go

Actions v

Descriptor	Number Of Errors
ACQDATE	44
BREDCODE	2
COLLDATE	4
COLLSRC	21
DONORCODE	21
ELEVATION	2
GENUS	1
LATITUDE	8
LONGITUDE	22
MLSSTAT	1
ORIGCTY	1
REMARKS	6
STORAGE	3

1 - 13

Deletion candidates

The number of accessions contained in an NI dataset may vary due to different reasons. For example, accessions could be removed from a certain genebank or the accession identifiers (unique combination of INSTCODE, GENUS, ACCENUMB) may change. With the new update mechanism, **no** accessions will be deleted automatically from EURISCO.

As a consequence of the new update mechanism, National Focal Points will explicitly have to name accessions to be deleted from the system.

In order to support this process, during the data integrity checks your new dataset was automatically compared with the existing dataset in EURISCO. This report provides an overview of accessions, which no longer exist in the new dataset, grouped by holding institution.

However, this list can only be a hint, which accessions could be candidates for deletion from EURISCO, and needs to be checked. Especially if only a part of the NI dataset should be updated, this list may contain many false positive entries.

Please send the checked list to weise@ipk-gatersleben.de.

Instcode	No Of Accessions
DEU005	36
DEU011	14
DEU021	25
DEU022	946
DEU043	8

1 - 5

Taxonomy check results

This dataset contains **732** distinct taxonomic names, which have been checked for correctness against the GRIN Taxonomy as well as against the Mansfeld's World Database of Agricultural and Horticultural Crops. The check results can be found [here](#).

Data integrity checks - example

- Wrong date format

✔
✔
●
●

Integrity Check

Nicode	Instcode	Genus	Accenumb	Descriptor	Error Description
AUT	AUT047	Brassica	HBLFAG-Bra-0019	ACQDATE	Line 2195: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0020	ACQDATE	Line 2196: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0021	ACQDATE	Line 2197: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0022	ACQDATE	Line 2198: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0023	ACQDATE	Line 2199: ACQDATE 2002--- invalid.
AUT	AUT047	Beta	HBLFAG-Bet-0005	ACQDATE	Line 2176: ACQDATE 2002--- invalid.
AUT	AUT047	Allium	HBLFAG-AII-0006	ACQDATE	Line 2165: ACQDATE 2002--- invalid.
AUT	AUT047	Allium	HBLFAG-AII-0007	ACQDATE	Line 2166: ACQDATE 2002--- invalid.
AUT	AUT047	Allium	HBLFAG-AII-0008	ACQDATE	Line 2167: ACQDATE 2002--- invalid.
AUT	AUT047	Beta	HBLFAG-Bet-0003	ACQDATE	Line 2174: ACQDATE 2002--- invalid.

1 - 10 >

Data integrity checks - example

- Invalid or multiple donor codes

Integrity Check

Nicode	Instcode	Genus	Accenumb	Descriptor	Error Description
AUT	AUT046	Triticum	ARCHE-WE267	DONORCODE	Line 650: DONORCODE FAO031 invalid.
AUT	AUT046	Triticum	ARCHE-WE268	DONORCODE	Line 651: DONORCODE FAO031 invalid.
AUT	AUT046	Triticum	ARCHE-WE269	DONORCODE	Line 652: DONORCODE FAO031 invalid.
AUT	AUT007	Camelina	BOKU-CU113	DONORCODE	Line 809: DONORCODE ROU006 invalid.
AUT	AUT001		BVAL-901795	DONORCODE	Line 11727: DONORCODE 21 invalid.
AUT	AUT001	Triticum	BVAL-256011	DONORCODE	Line 8854: DONORCODE DEU146 / DEU235 invalid.
AUT	AUT001	Triticum	BVAL-256013	DONORCODE	Line 8856: DONORCODE DEU146 / USA126 invalid.
AUT	AUT001	Digitaria	BVAL-903441	DONORCODE	Line 11173: DONORCODE AUT013/DEU146 invalid.
AUT	AUT001	Chenopodium	BVAL-903215	DONORCODE	Line 11073: DONORCODE DEU146 / DEU032 invalid.
AUT	AUT001	Chenopodium	BVAL-903216	DONORCODE	Line 11074: DONORCODE DEU146 / SVN006 invalid.

1 - 10 >

Incremental updates

- Why incremental updates?
 - In the past: Full replacement
 - Delete whole dataset + reimport data afterwards
 - Even if only a couple of rows have been modified
 - Not possible to update parts of data (e.g. single genebank collection)
 - Thus: From full replacement to real update
 - Only incremental data needs to be updated
 - Necessary: Unique identifiers
 - So far: Combination of NICODE, INSTCODE, ACCENUMB and GENUS
 - Now: DOIs (e.g. by using infrastructure of ITPGRFA)
 - Important for managing C&E data
 - Cannot exist without passport data

Incremental updates

- Deletion candidates
 - Check of new dataset against existing data
 - List of accessions not contained in the new dataset
 - Not deleted automatically
- Just a hint
 - Needs to be checked by NFP
 - False positive hits in case of partial update!!!

Deletion candidates

The number of accessions contained in an NI dataset may vary due to different reasons. For example, accessions could be removed from a certain genebank or the accession identifiers (unique combination of INSTCODE, GENUS, ACCENUMB) may change. With the new update mechanism, **no** accessions will be deleted automatically from EURISCO.

As a consequence of the new update mechanism, National Focal Points will explicitly have to name accessions to be deleted from the system.

In order to support this process, during the data integrity checks your new dataset was automatically compared with the existing dataset in EURISCO. This report provides an overview of accessions, which no longer exist in the new dataset, grouped by holding institution.

However, this list can only be a hint, which accessions could be candidates for deletion from EURISCO, and needs to be checked. Especially if only a part of the NI dataset should be updated, this list may contain many false positive entries.

Please send the checked list to weise@ipk-gatersleben.de.

Instcode	No Of Accessions
DEU005	36
DEU011	14
DEU021	25
DEU022	946
DEU043	8

1 - 5

Nicode	Instcode	Genus	Accenumb
DEU	DEU146	PISUM	PIS 1643
DEU	DEU146	PISUM	PIS 1644
DEU	DEU146	PISUM	PIS 1645
DEU	DEU146	PISUM	PIS 1646
DEU	DEU146	PISUM	PIS 1647
DEU	DEU146	PISUM	PIS 1648
DEU	DEU146	PISUM	PIS 1649
DEU	DEU146	PISUM	PIS 165
DEU	DEU146	PISUM	PIS 1650
DEU	DEU146	PISUM	PIS 1651

1 - 10

AEGIS status checks

- Automatic check for de-flagging of AEGIS accessions
 - Works in case of incremental updates too
 - Reason for change needs to be documented and will be sent to the ECPGR Secretariat

Changes of AEGIS status

During the data integrity checks, the AEGIS status of the accessions of the new dataset was compared with the current dataset in EURISCO. The report below shows accessions existing in EURISCO, which were labelled as part of AEGIS. These accessions are either missing from the new dataset or the AEGISSTAT value changed.

In accordance with the AEGIS principles, please enter a reason for removing these accessions from AEGIS. This information will be sent to the ECPGR Secretariat automatically.

Changes of AEGIS status

Instcode	Genus	Accenumb	Aegisstat Old	Aegisstat New	Change Reason	Approved At	Approve Link
DEU146	VIGNA	VIG 134	1	0	<input type="text"/>		approve
DEU146	VIGNA	VIG 135	1		<input type="text"/>		approve

1 - 2

Final decision



Final Decision

Fourth Step : Publish/Discard new data

After you have reviewed the errors which occurred during the data integrity checks, the fourth step of importing new or modified data into EURISCO is now to decide either to publish the new data to the EURISCO web frontend or to discard the imported data. In the latter case, the whole update procedure should be repeated with a debugged data set.



Go

Actions X

National
Inventory

File name

DEU

49227975754531859/test_f



Decision on update

Final decision

Your uploaded file has been checked for integrity and can now be used to update the data of your National Inventory in EURISCO. Only accessions listed in your file will be updated. All other accessions of your National Inventory in EURISCO, which are not covered by the input file, will remain untouched in EURISCO.

The final update will run as a batch job in the background.

Update EURISCO data

Discard data

Next steps (background process)

- Updated dataset will be applied to EURISCO stage schema
- EURISCO stage will be synchronised to the EURISCO web schema (time lag!)
 - Not in main business hours
 - Rebuild of materialised views
 - Creation of new full dump (MS Access)
 - News message on EURISCO webpage

Reminder: Migration to v2.1 of MCPD

- Until end of 2017, the data exchange format was based on MCPD v1 (December 2001)
 - MCPD v1
 - 8 additional, EURISCO-specific descriptors
- In the meantime, evolution of MCPD to v2.1 (Dec 2015)
 - Adaptation of the EURISCO exchange format
 - Harmonisation with MCPD 2.1
 - 4 additional descriptors

Reminder: Migration to v2.1 of MCPD

- New descriptors
 - PUID
 - Persistent unique identifier, e.g. DOI
 - COLLINSTADDRESS
 - Address of collecting institute
 - COLLMISSID
 - Identifier of collecting mission
 - DECLATITUDE
 - Latitude in decimal degrees
 - DECLONGITUDE
 - Longitude in decimal degrees
 - COORDUNCERT
 - Uncertainty of coordinates in metres
 - COORDDATUM
 - Geodetic datum or reference system, e.g. WGS84
 - GEOREFMETH
 - Referencing methods, e.g. GPS
 - HISTORIC
 - Accession maintenance status

Reminder: Migration to v2.1 of MCPD

- Modified descriptors
 - COLLCODE: multiple values allowed
 - DUPLSITE: multiple values allowed
 - BREDCODE: multiple values allowed

 - COLLNAME: replaces COLLDESCR
 - BREDNAME: replaces BREDDESCR
 - DONORNAME: replaces DONORDESCR
 - DUPLINSTNAME: replaces DUPLDESCR