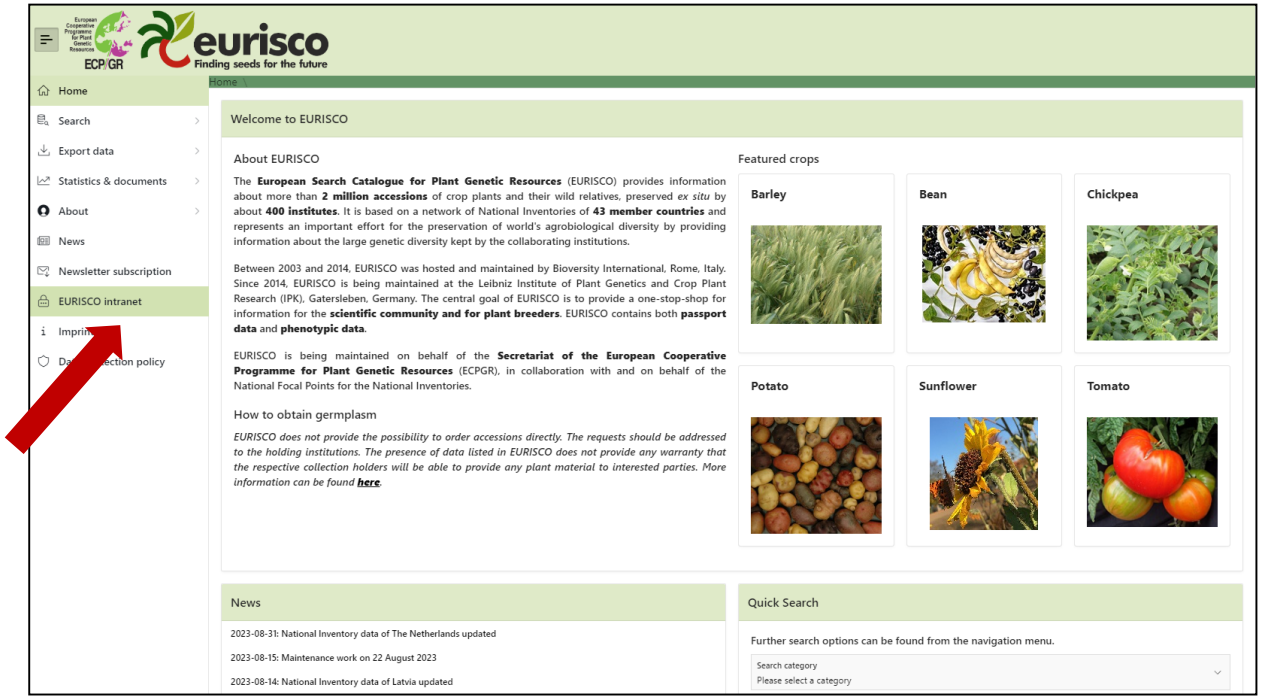
Stephan Weise

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



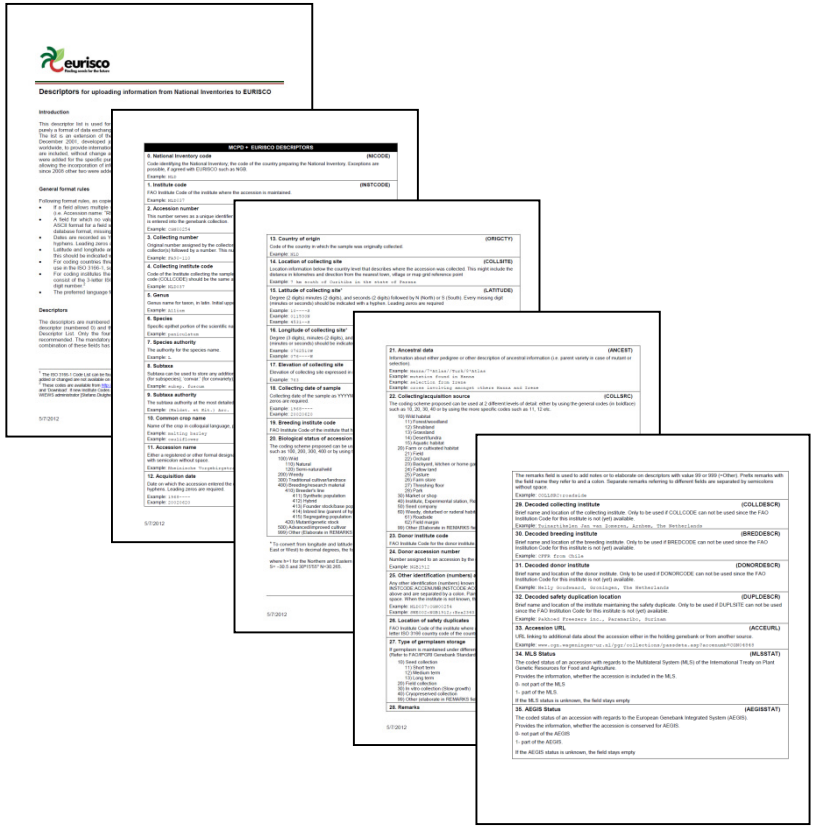
EURISCO intranet

Step 1
Upload file

Step 2
Import File

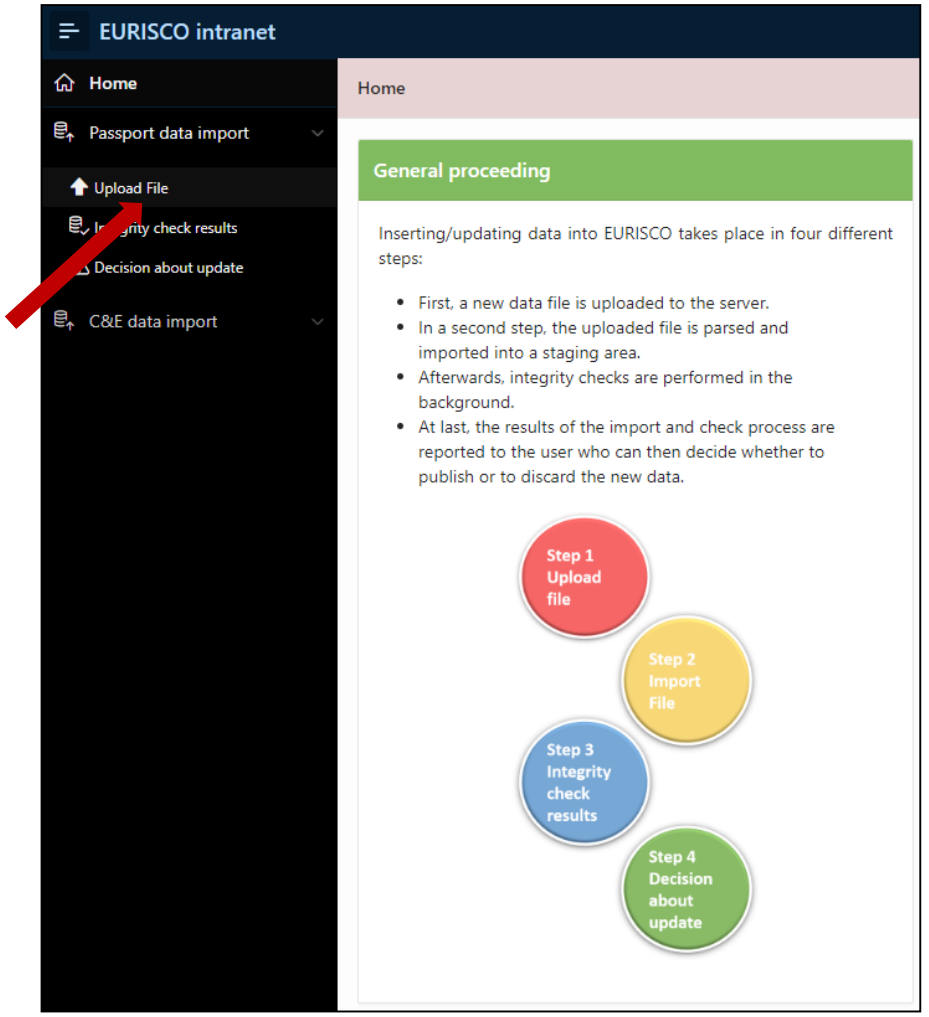
Step 3
Integrity check results

Step 4
Decision about update



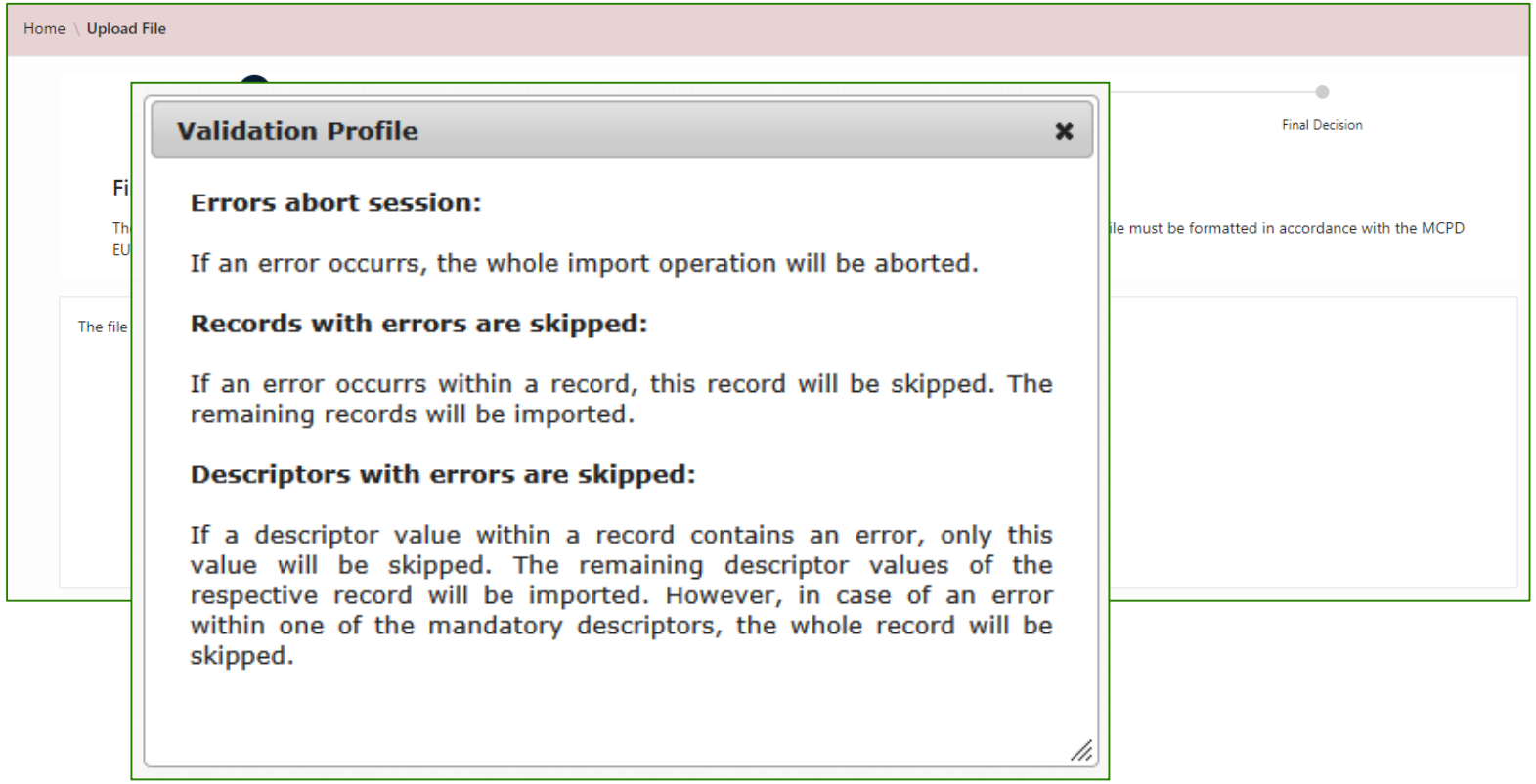
File upload

- 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”



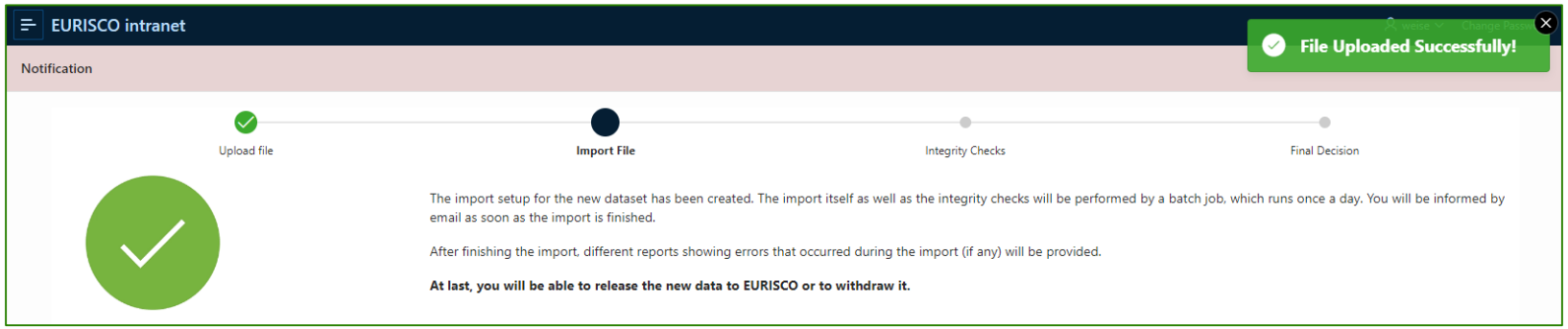
File upload

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



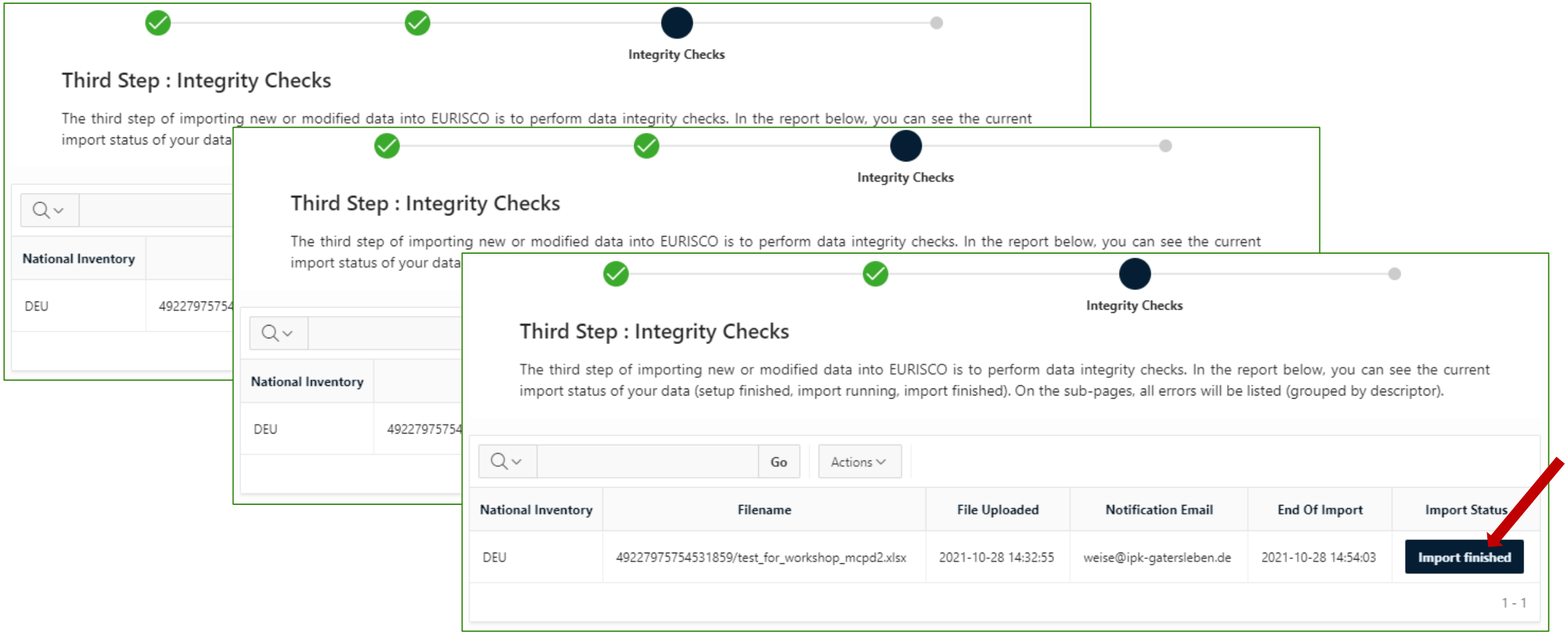
File import

- 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

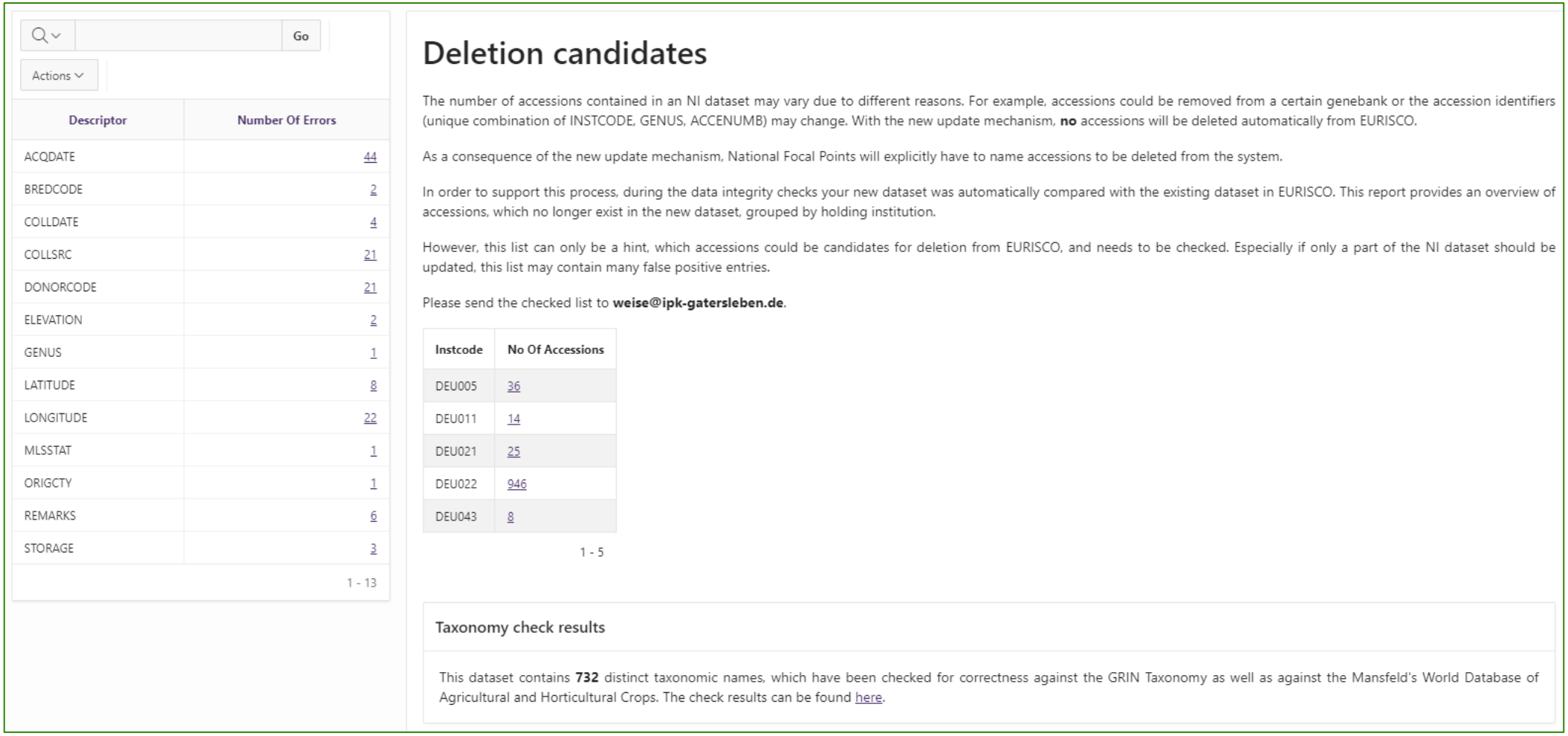


No input needed from your side

Data integrity checks



Data integrity checks



Go

Actions

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

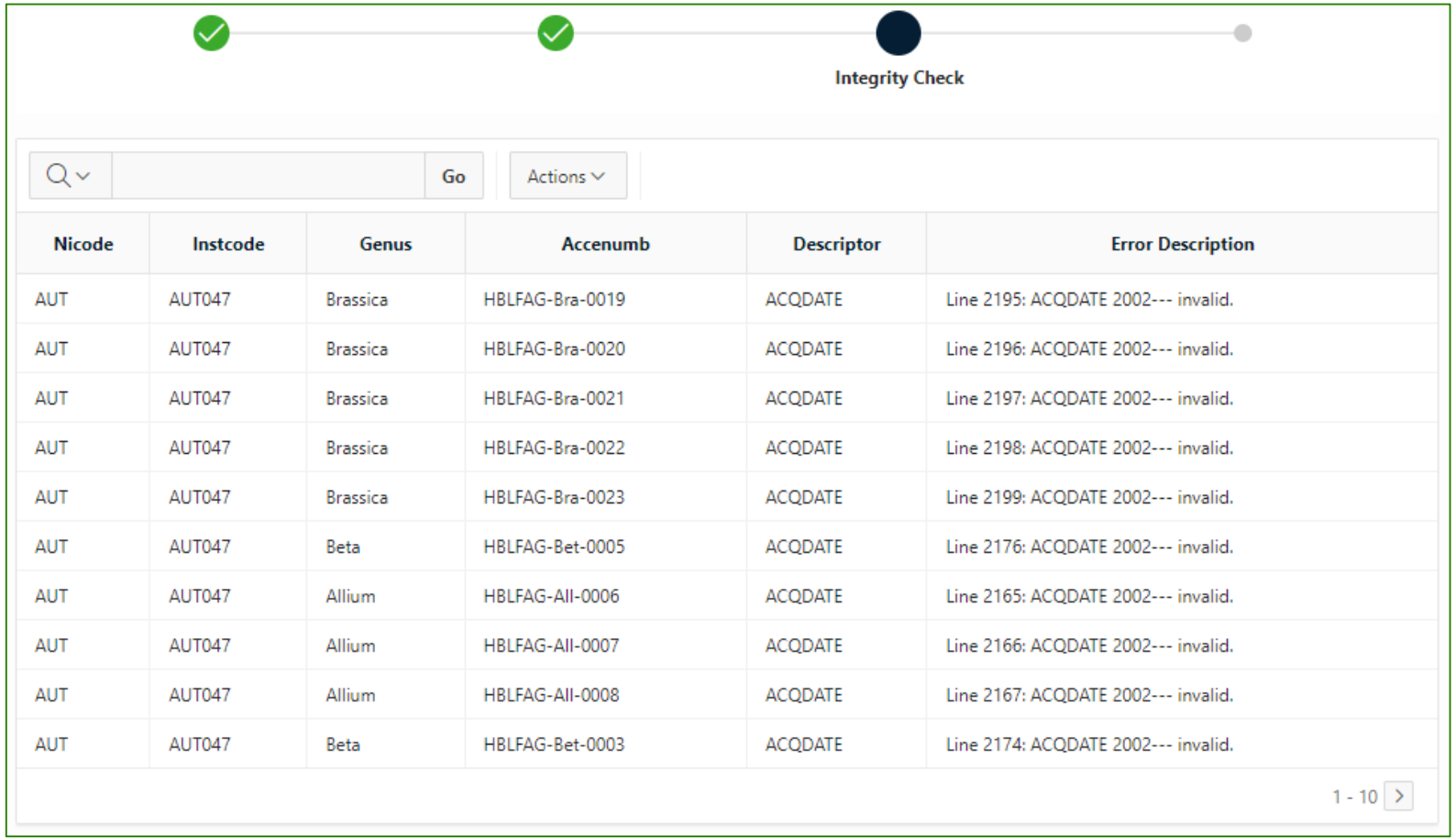
1 - 5

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

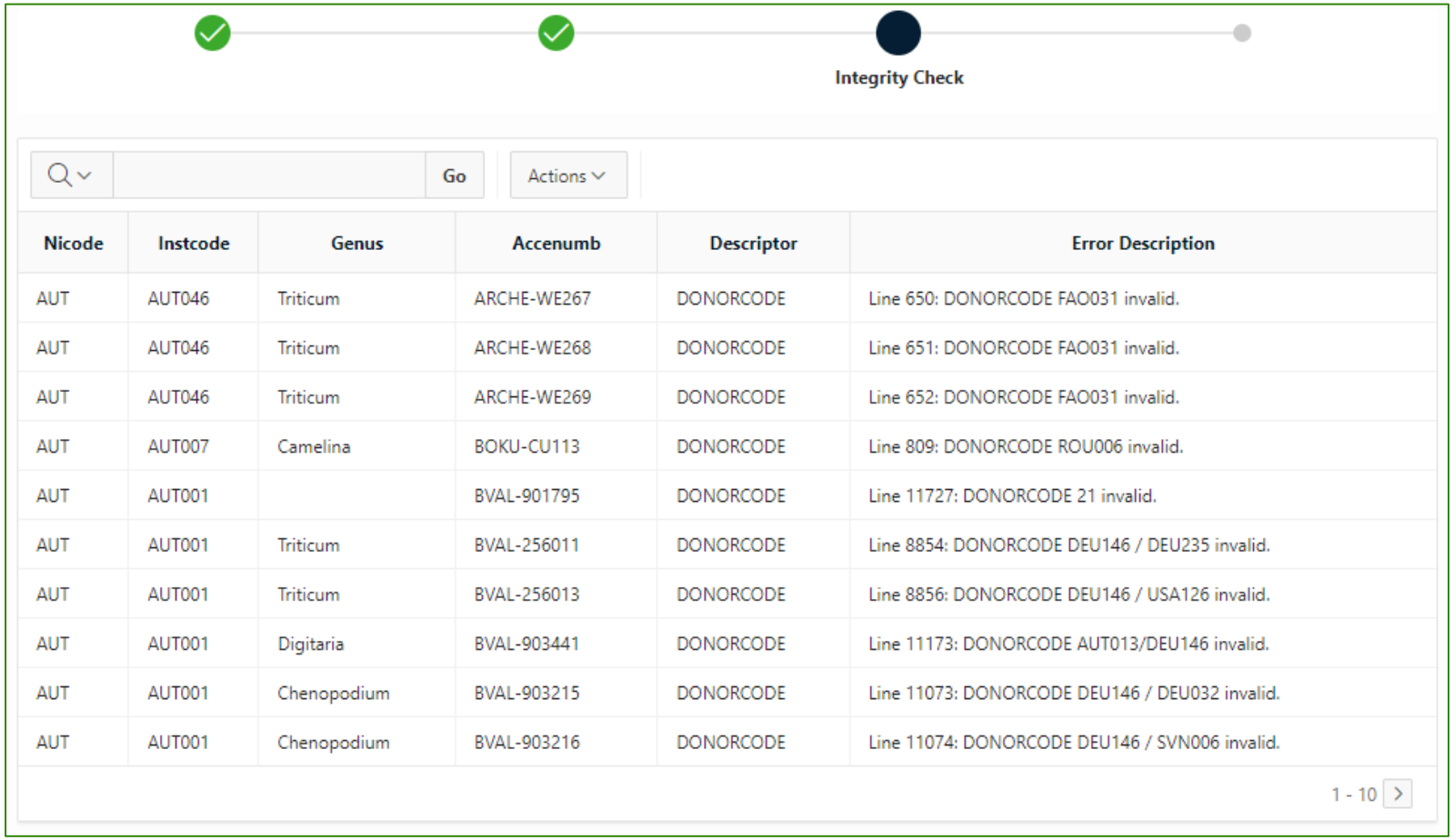


Nicode	Instcode	Genus	Accenumb	Descriptor	Error Description
AUT	AUT047	Brassica	HBLFAG-Bra-0019	ACQDATE	Line 2195: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0020	ACQDATE	Line 2196: ACQDATE 2002--- invalid.
AUT	AUT047	Brassica	HBLFAG-Bra-0021	ACQDATE	Line 2197: ACQDATE 2002--- invalid.
AUT	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-All-0006	ACQDATE	Line 2165: ACQDATE 2002--- invalid.
AUT	AUT047	Allium	HBLFAG-All-0007	ACQDATE	Line 2166: ACQDATE 2002--- invalid.
AUT	AUT047	Allium	HBLFAG-All-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



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.

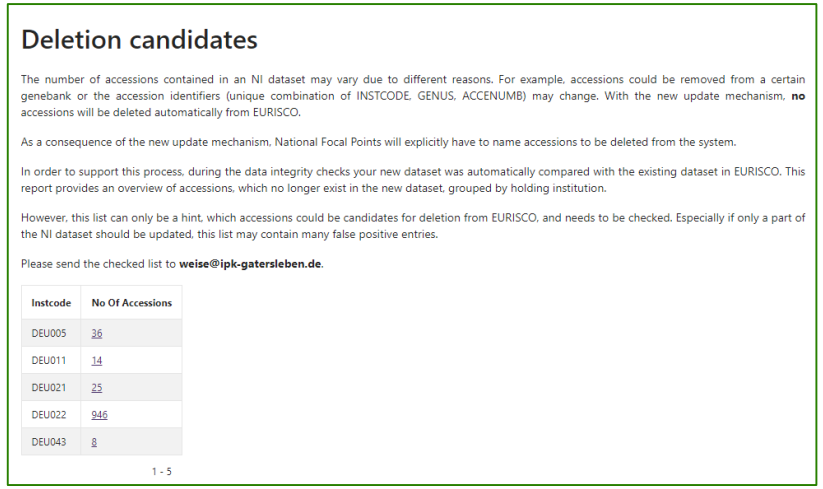
1 - 10

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.

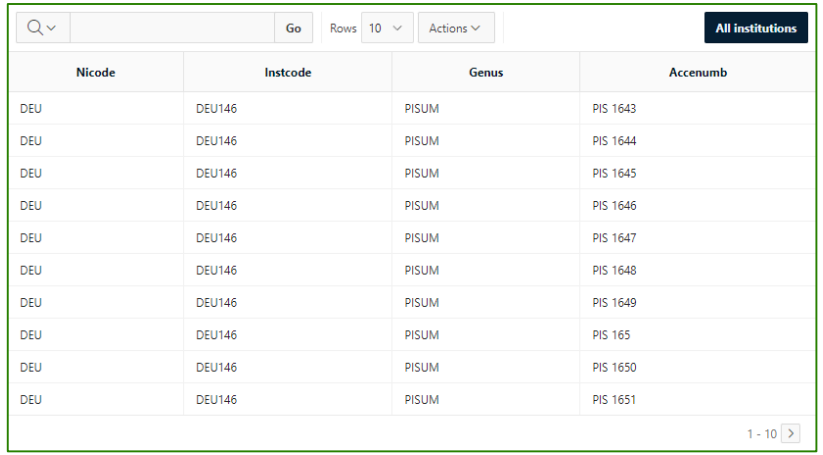
In order to support this process, during the data integrity checks your new dataset was automatically compared with the existing dataset in EURISCO. This report provides an overview of accessions, which no longer exist in the new dataset, grouped by holding institution.

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

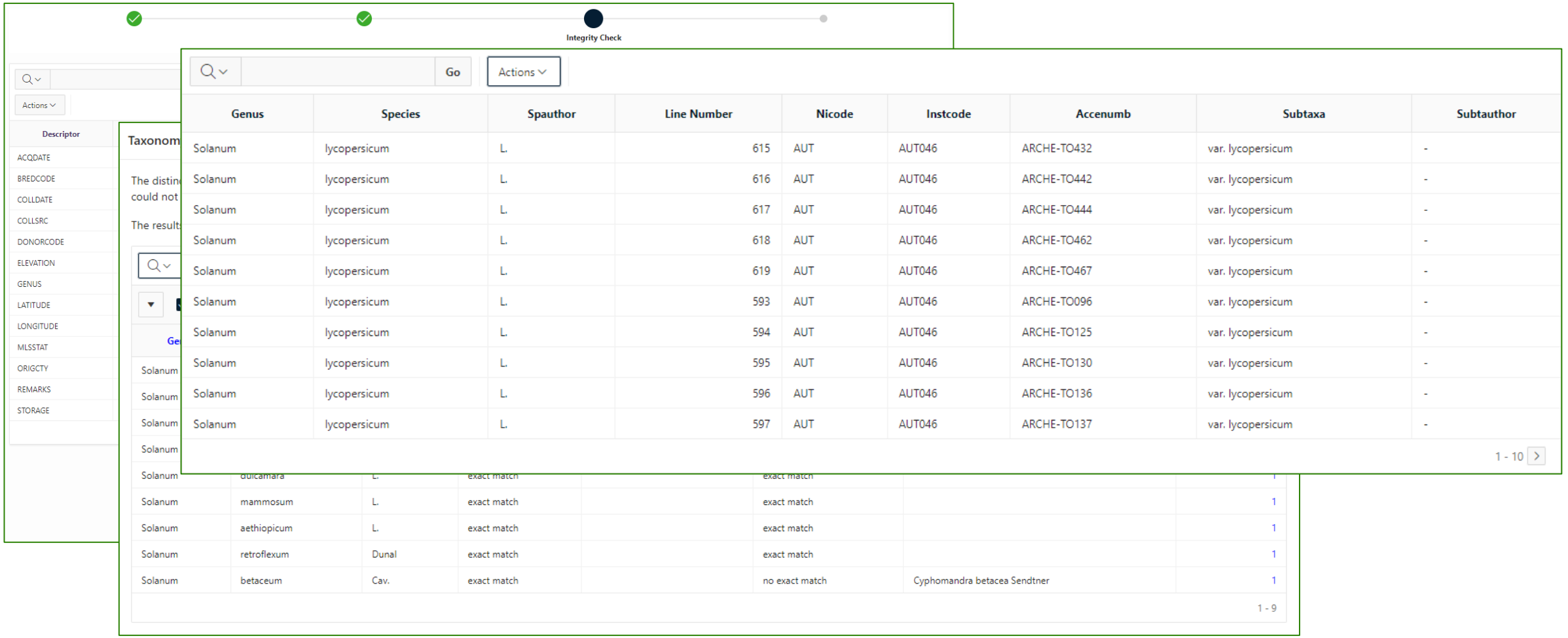


Go | Rows 10 | Actions | All institutions

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

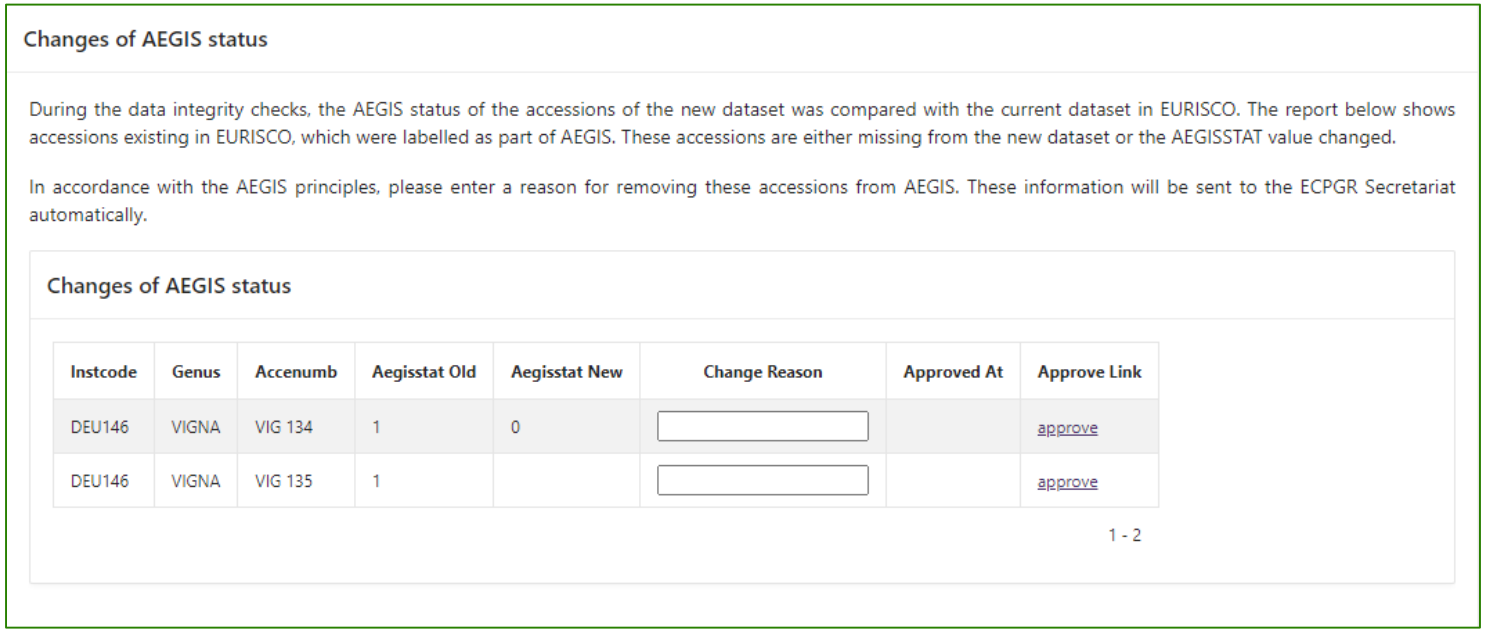
1 - 10

Taxonomy checks



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.

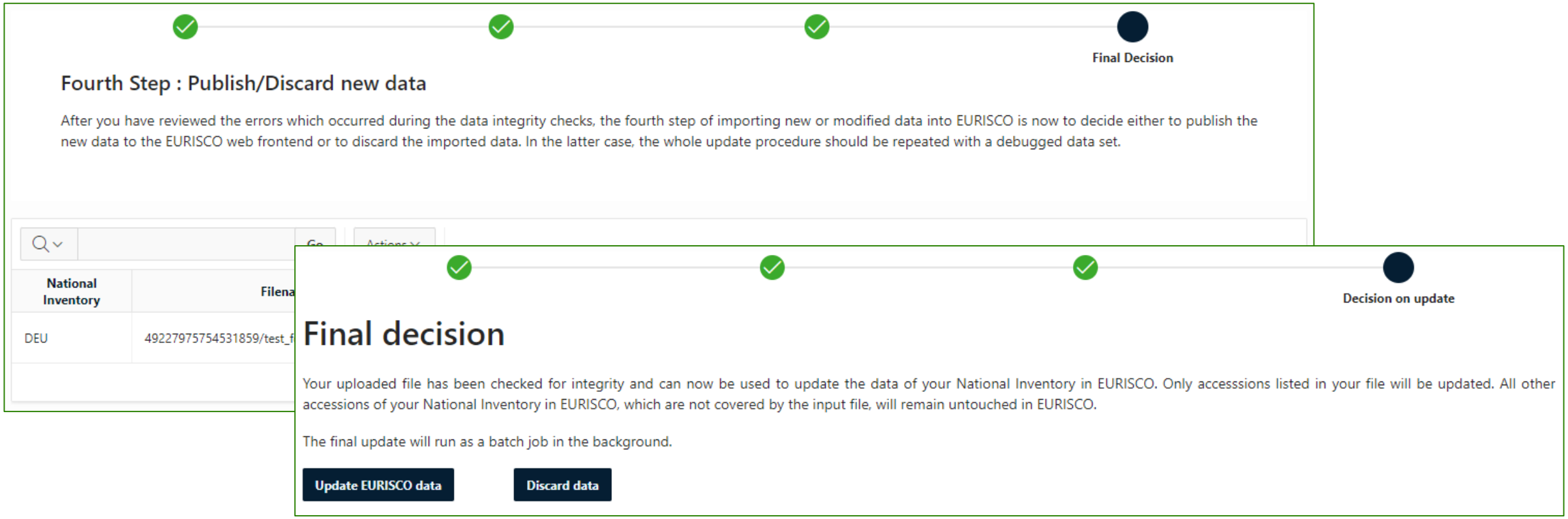
In accordance with the AEGIS principles, please enter a reason for removing these accessions from AEGIS. These information will be sent to the ECPGR Secretariat automatically.

Changes of AEGIS status

Instcode	Genus	Accenumb	Aegisstat Old	Aegisstat New	Change Reason	Approved At	Approve Link
DEU146	VIGNA	VIG 134	1	0			approve
DEU146	VIGNA	VIG 135	1				approve

1 - 2

Final decision



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