Passport data updates

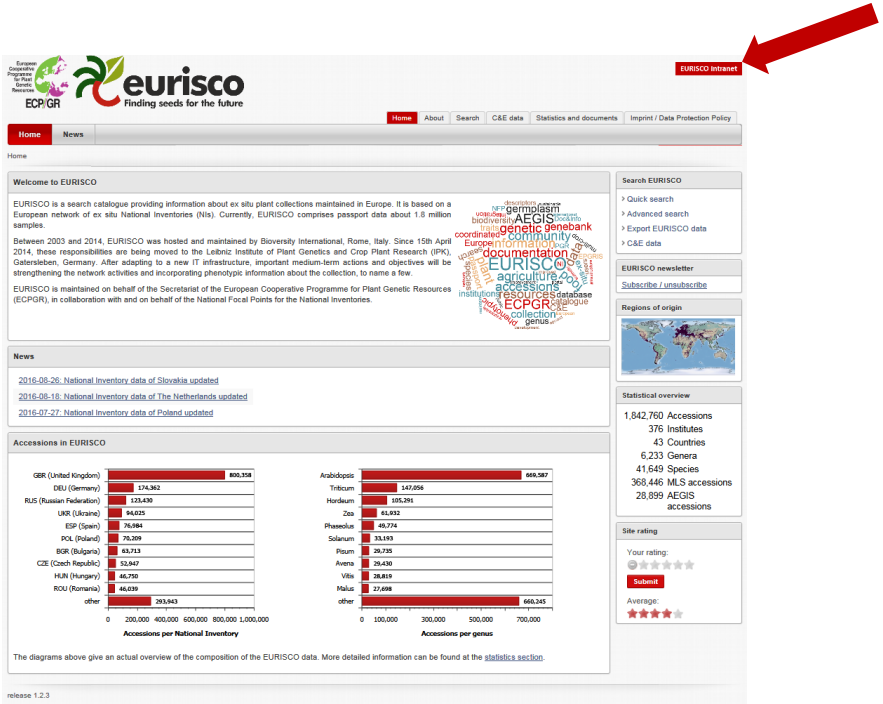
Presentation & discussion

EURISCO training workshop, 12th to 14th October 2016, Angers, France

Stephan Weise
13 October 2016

EURISCO intranet I

- Development of new import component for NIs
  - Web interface with Oracle APEX
  - PL/SQL packages for uploading, checking and updating data
  - Implementation of incremental updates



EURISCO intranet II

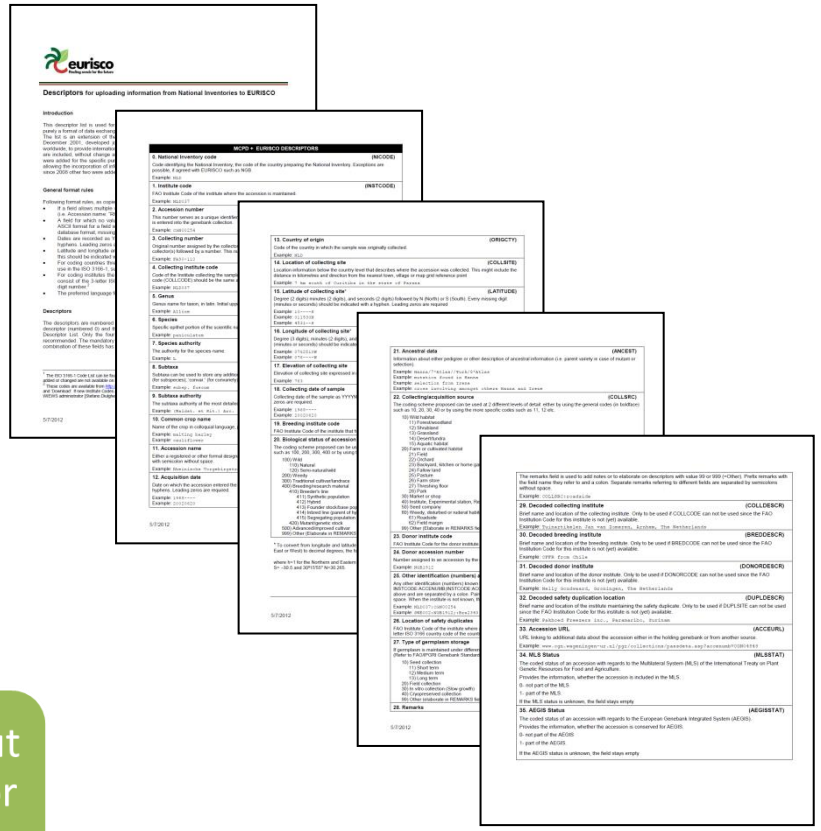
Upload file to server

Import file into database

New Java-based importer

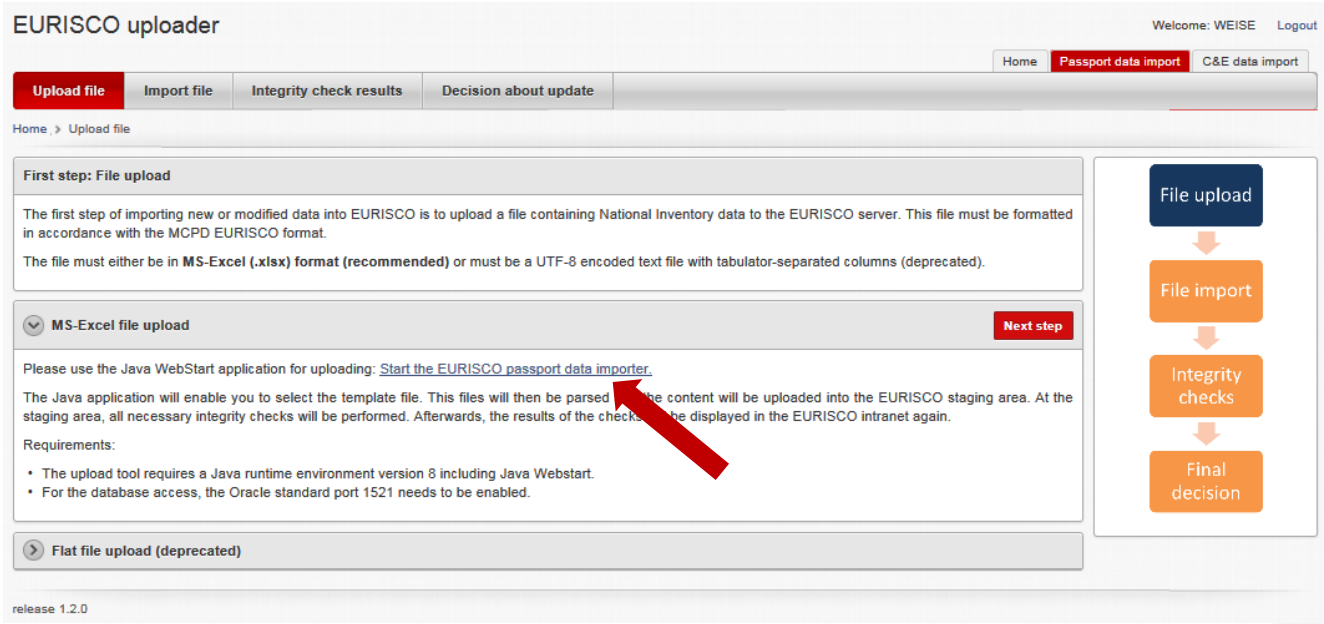
Perform data checks

Decide about publishing or withdrawel



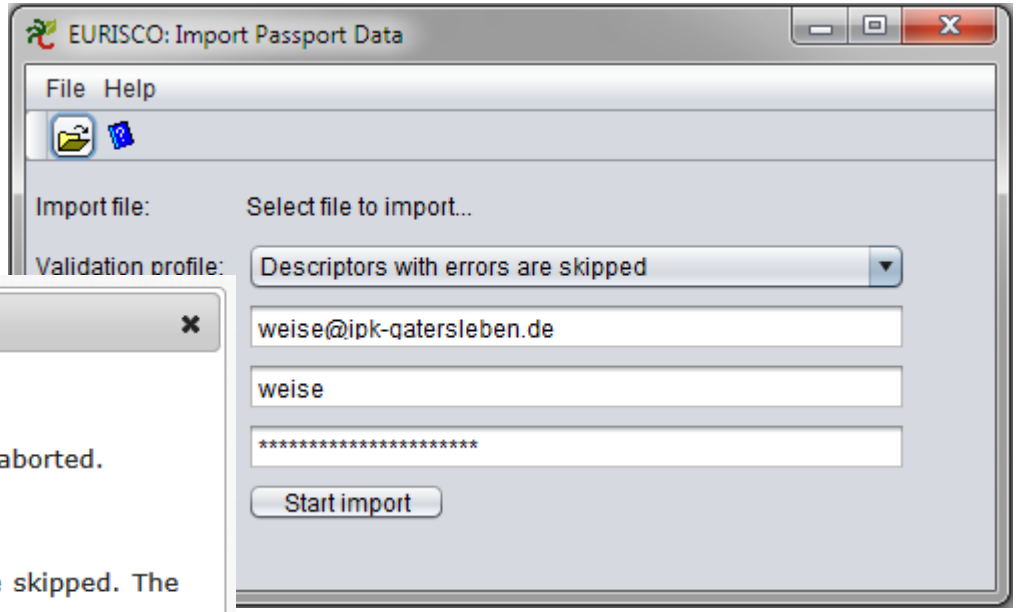
Java-based import I

- So far: Web-based upload
  - Upload via tab-separated text file, UTF-8 encoded
  - Often problems with columns separators + character encoding
- Now: Java-based Excel import
  - Summarises file upload + data import
  - User gets informed when the integrity checks are finished



Java-based import II

- JRE 1.8
- Java WS
- Oracle standard port 1521 enabled



Validation Profile

Errors abort session:

If an error occurrs, the whole import operation will be aborted.

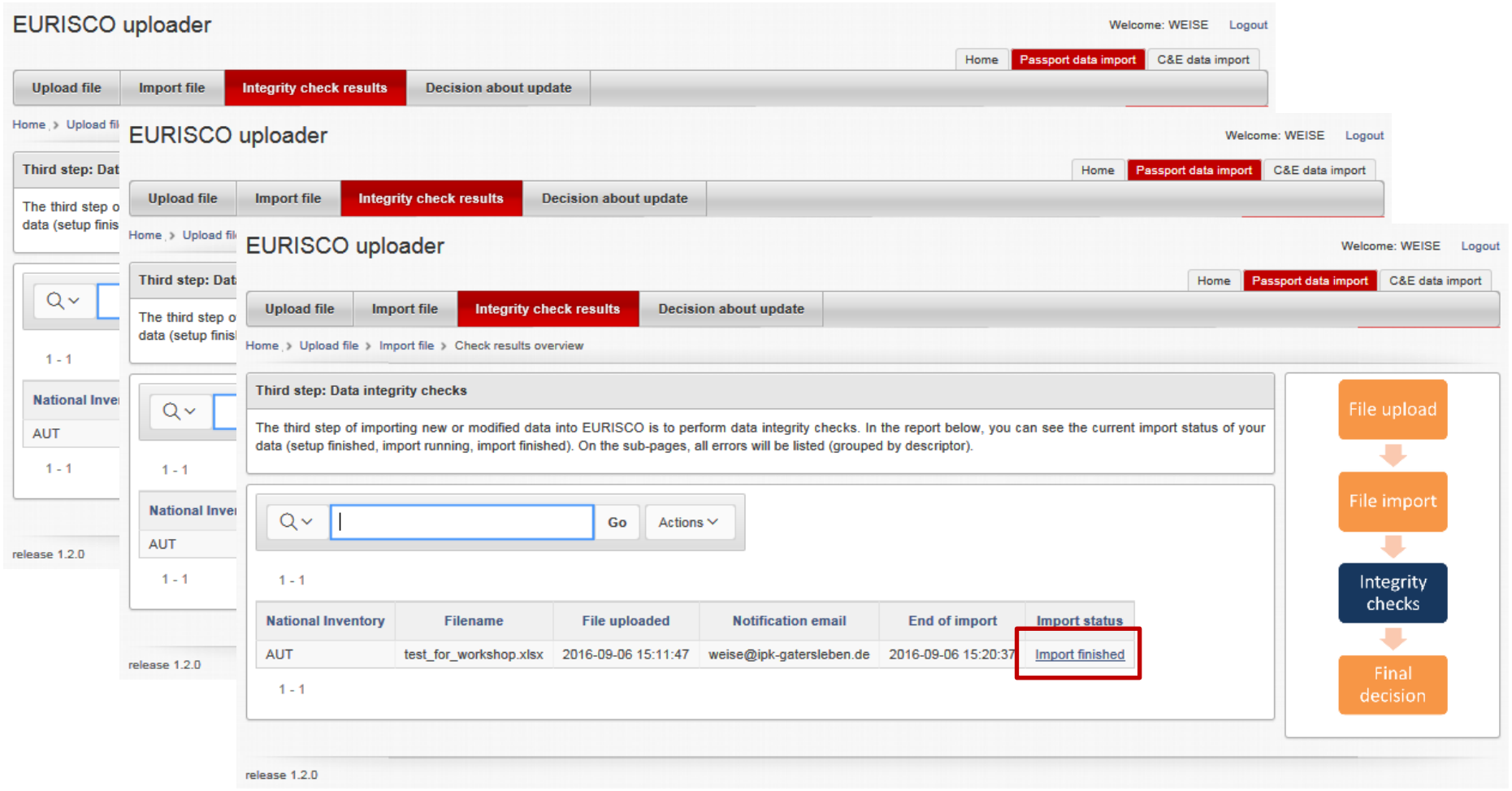
Records with errors are skipped:

If an error occurrs within a record, this record will be skipped. The remaining records will be imported.

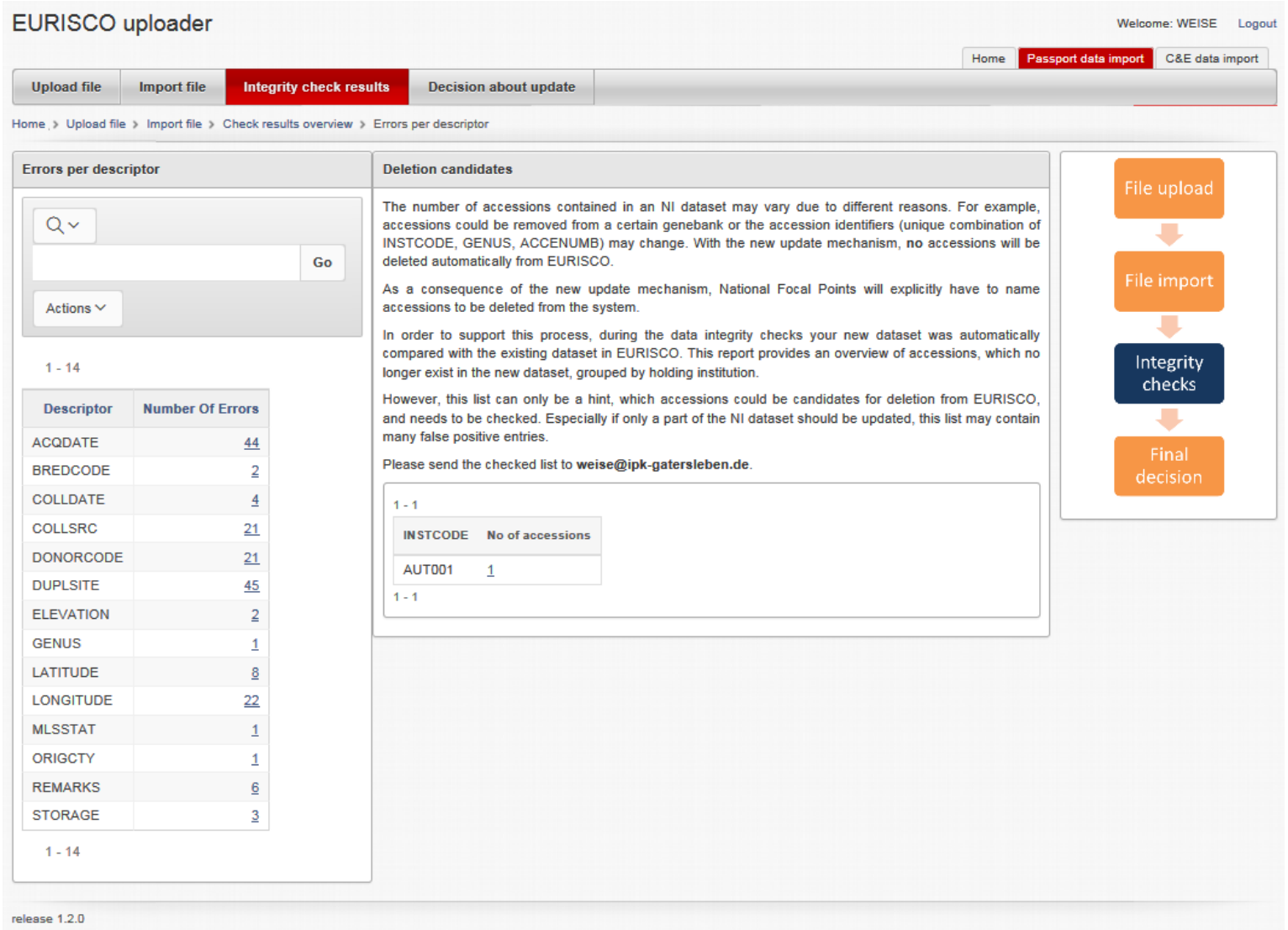
Descriptors with errors are skipped:

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.

Integrity checks I

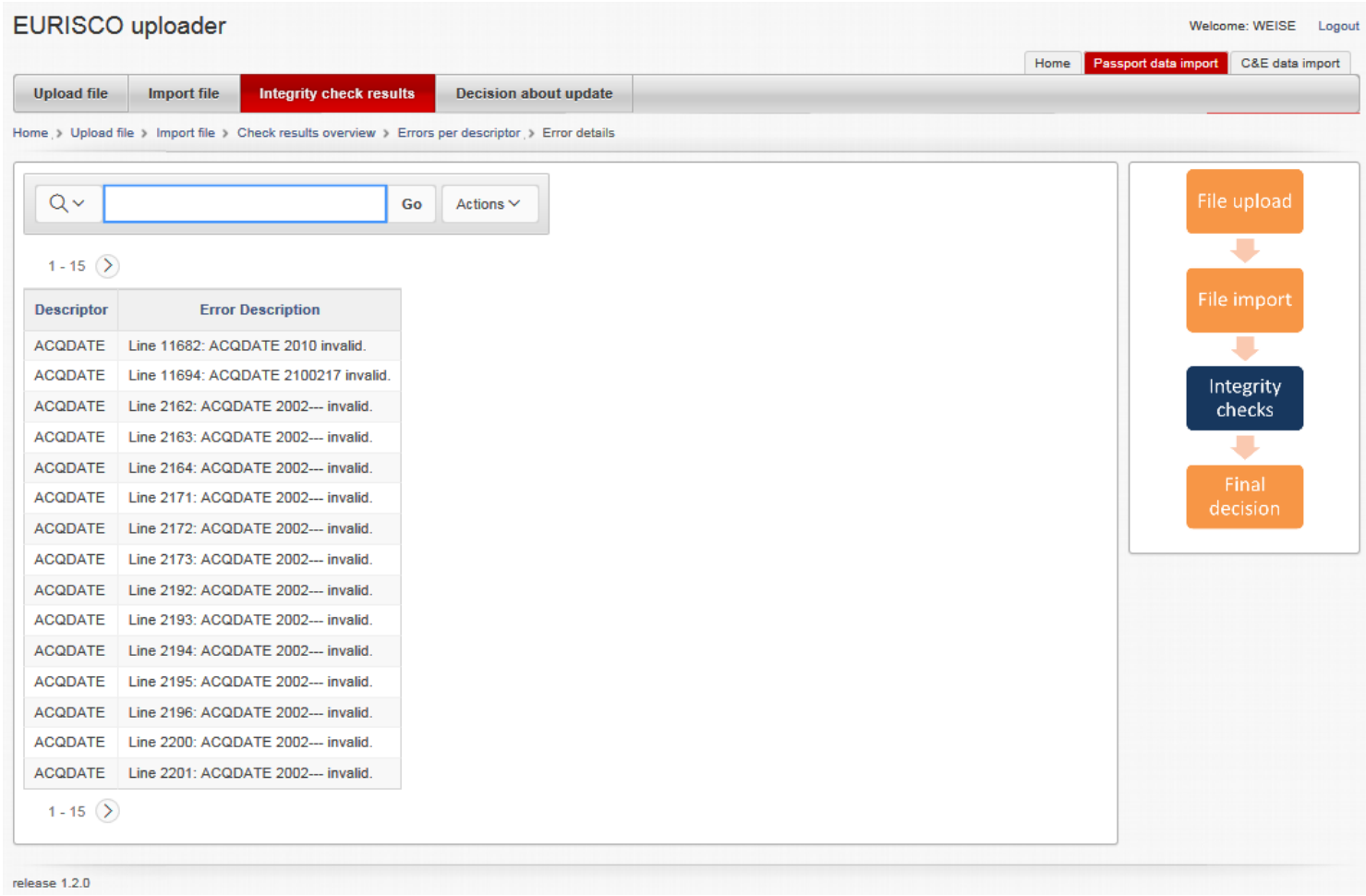


Integrity checks II – error report



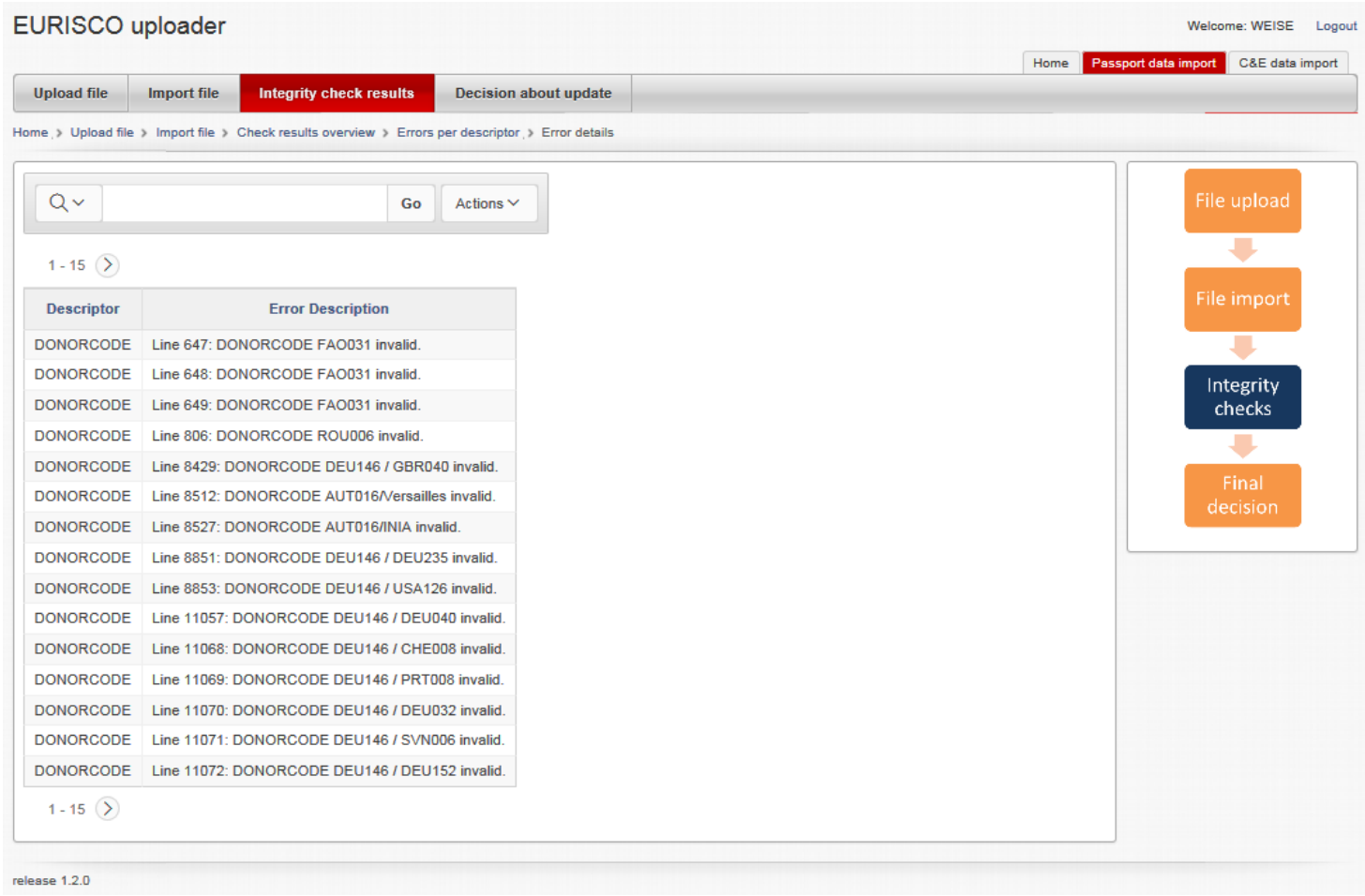
Integrity checks III - example

- Wrong date format



Integrity checks IV - example

- Invalid or multiple donor codes

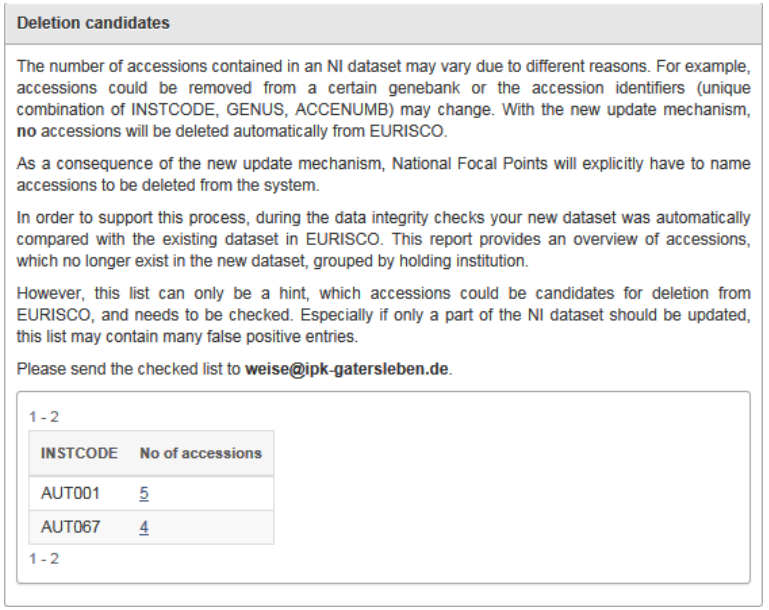


Incremental updates I

- Why incremental updates?
  - So far: 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)
  - That's why: From full replacement to real update
    - Only incremental data needs to be updated
    - Necessary: Unique identifiers
    - Currently: Combination of NICODE, INSTCODE, ACCENUMB and GENUS
  - Important for managing C&E data
    - Cannot exist without passport data

Incremental updates II

- Deletion candidates
  - Check of new dataset against existing data
  - List of accessions not contained in the new dataset
  - Not deleted automatically
  - 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.

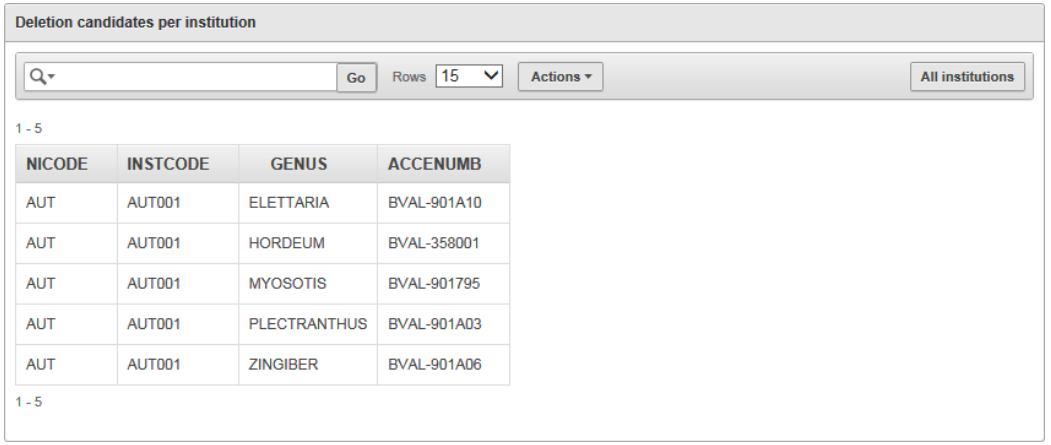
However, this list can only be a hint, which accessions could be candidates for deletion from EURISCO, and needs to be checked. Especially if only a part of the NI dataset should be updated, this list may contain many false positive entries.

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

1 - 2

INSTCODE	No of accessions
AUT001	5
AUT067	4

1 - 2



Deletion candidates per institution

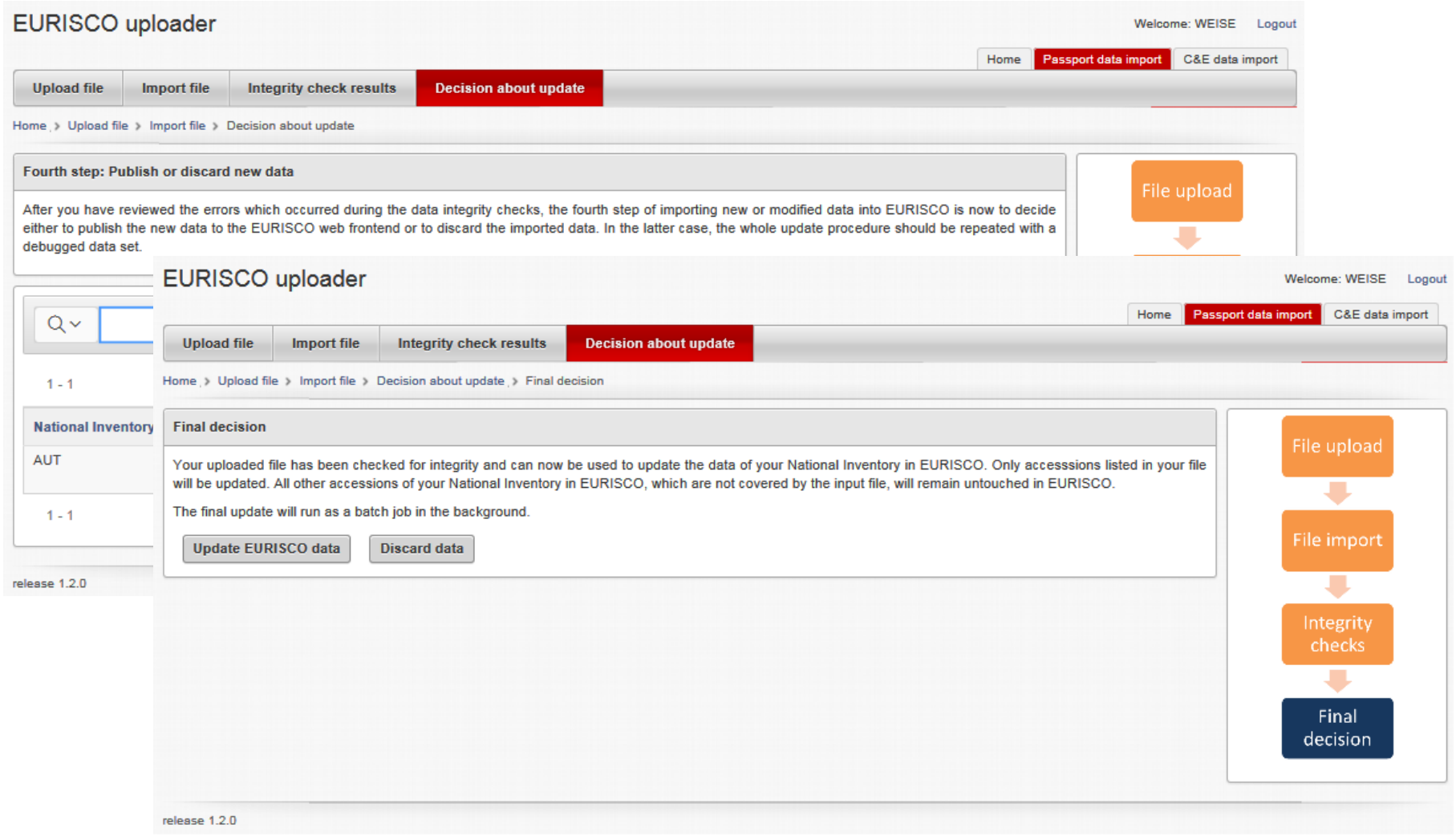
Go | Rows 15 | Actions | All institutions

1 - 5

NICODE	INSTCODE	GENUS	ACCENUMB
AUT	AUT001	ELETTARIA	BVAL-901A10
AUT	AUT001	HORDEUM	BVAL-358001
AUT	AUT001	MYOSOTIS	BVAL-901795
AUT	AUT001	PLECTRANTHUS	BVAL-901A03
AUT	AUT001	ZINGIBER	BVAL-901A06

1 - 5

Final decision



Next steps (in background)

- 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