EURISCO catalogue

Status quo & planned developments

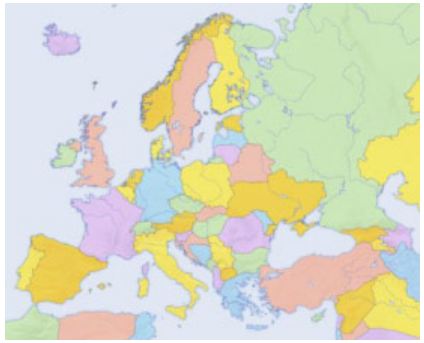
EURISCO training workshop, 19th to 21st May 2015, Tirana, Albania

Stephan Weise
19 May 2015

BACKGROUND

Introduction I

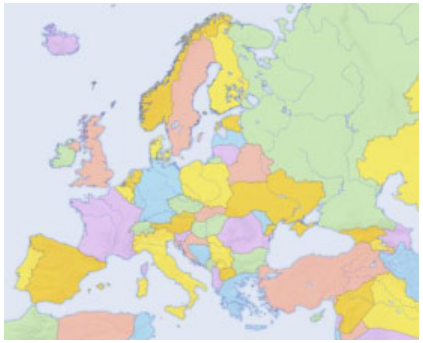
- Development of European information system for plant genetic resources
  - Started in 2001 (EU project EPGRIS)
  - EURISCO: Search catalogue for *ex situ* collections; available since 2003
  - National collections represented by National Inventories (NIs)
  - 43 countries involved (Nordic Countries → NGB)
  - Network of NFPs links NIs ↔ EURISCO



https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png

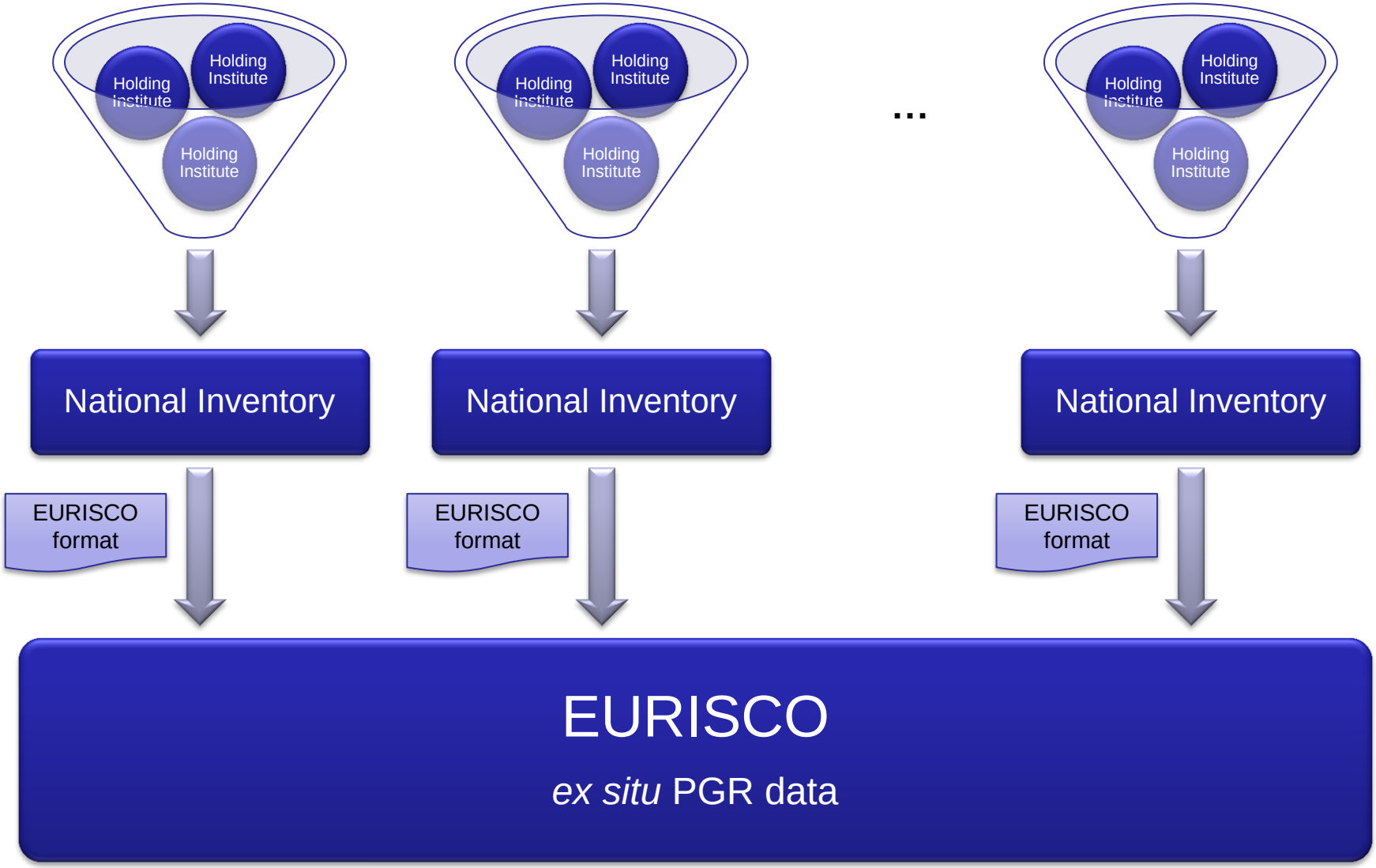
Introduction II

- Development of European information system for plant genetic resources (cont.)
  - Accession-level information system
  - Provides passport information about PGR maintained in Europe
  - Assists in meeting national obligations
    - FAO
    - CBD
    - ITPGR
    - (Nagoya ???)



https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png

Introduction III



Some statistics I

- 1,117,040 accessions
- 363 holding institutes
- 43 countries
- 5,657 genera
- 36,557 species (unique comb. genus + species)
- 325,340 MLS accessions
- 15,133 AEGIS accessions

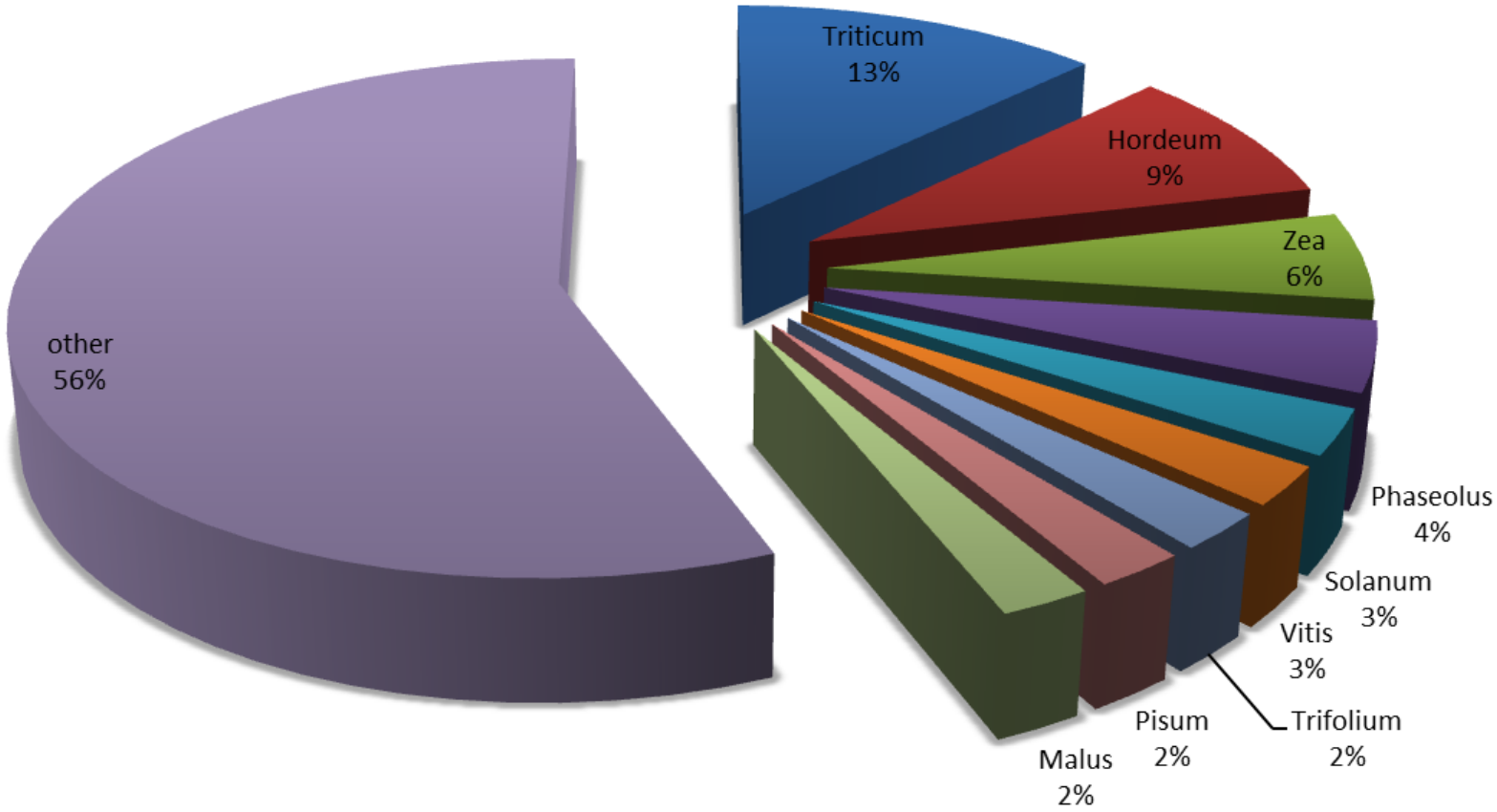
as of 2015-04-17

Some statistics II

- 712,731 accessions with collecting information
  - 87,132 different collecting sites
  - But only 172,910 accessions with coordinates (15% of all accessions)
- 518,820 accessions with donor information
- 879,347 accessions with country of origin
  - 23 different geographic regions

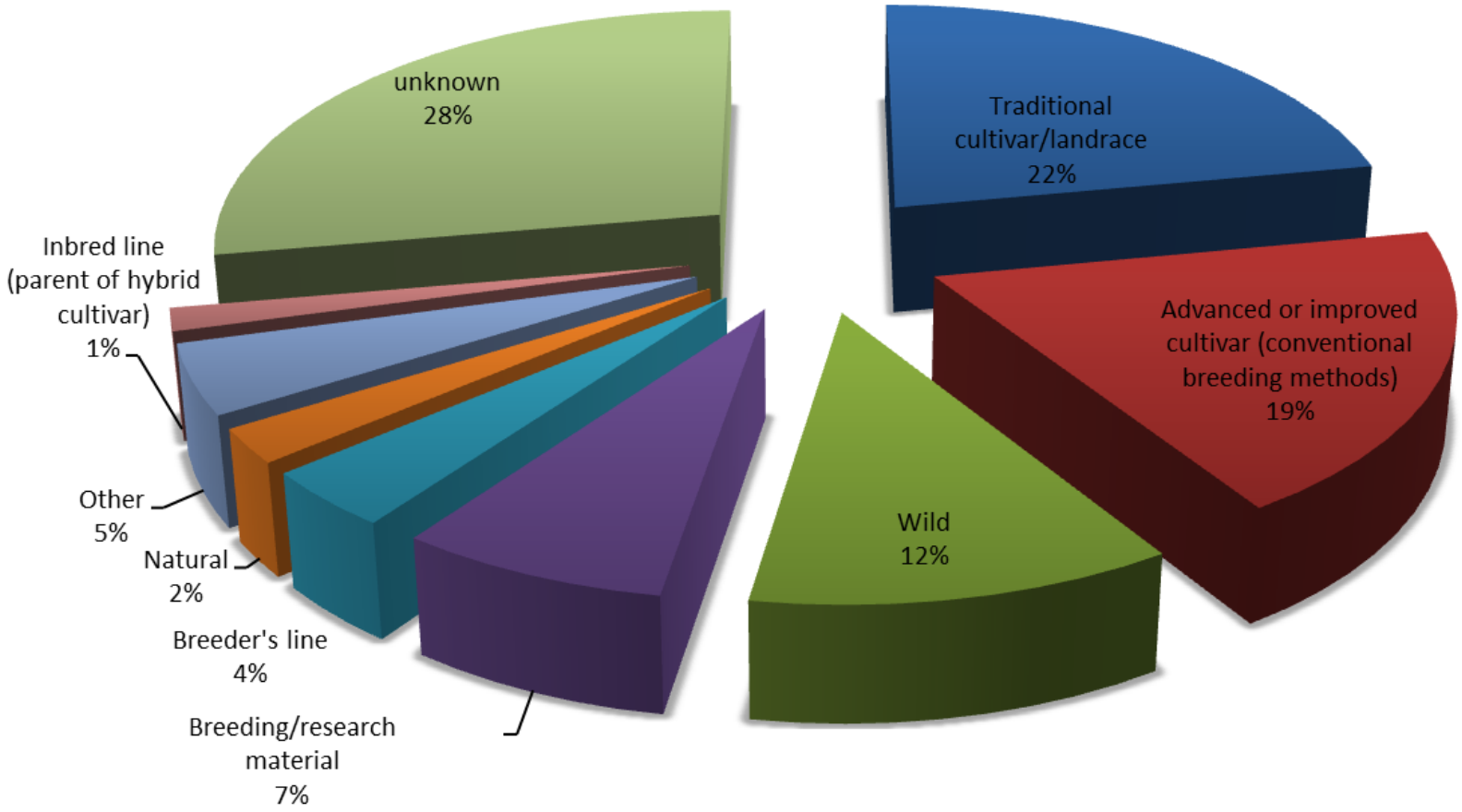
as of 2015-04-17

Genera



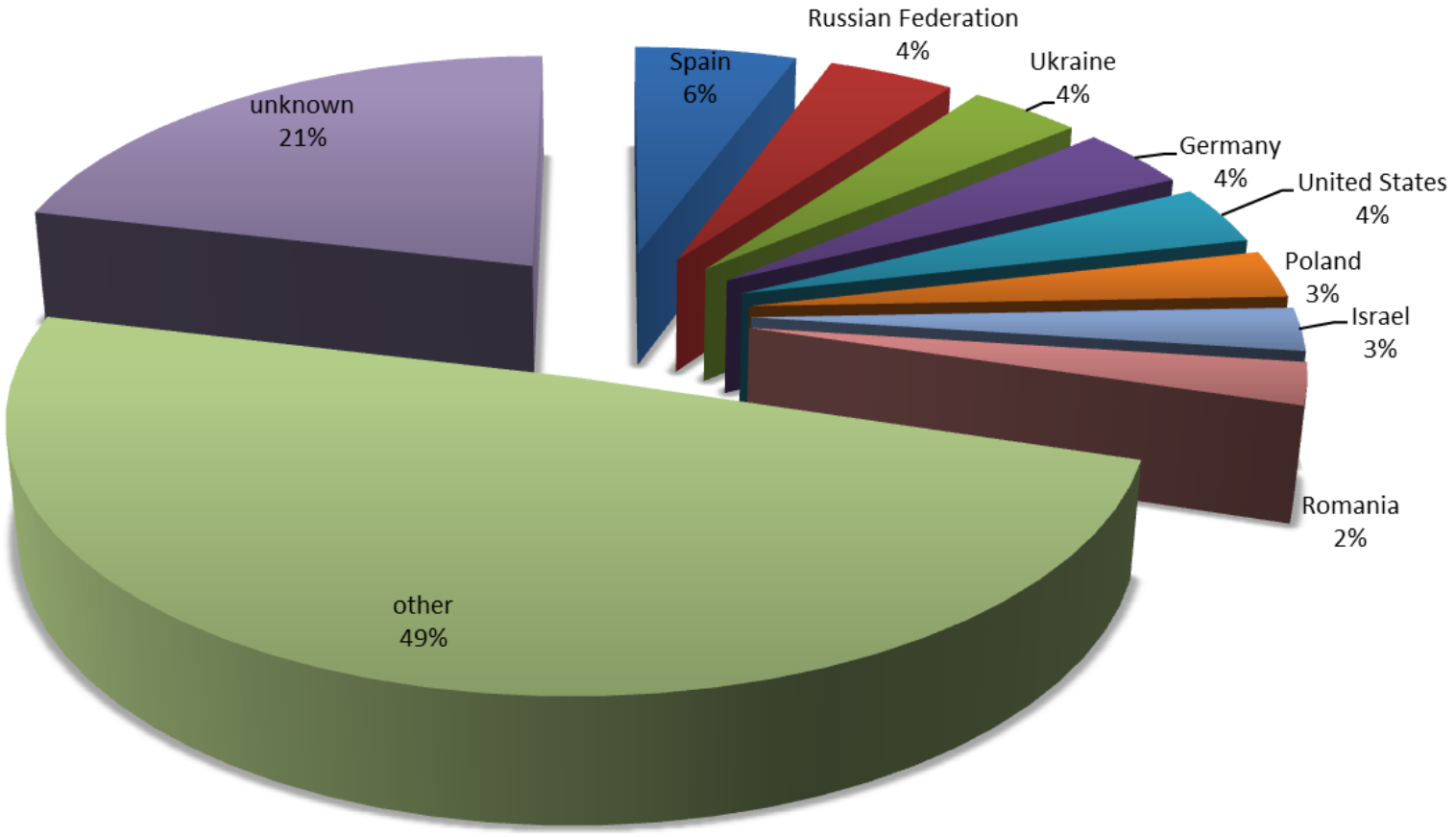
as of 2015-04-17

Biological status



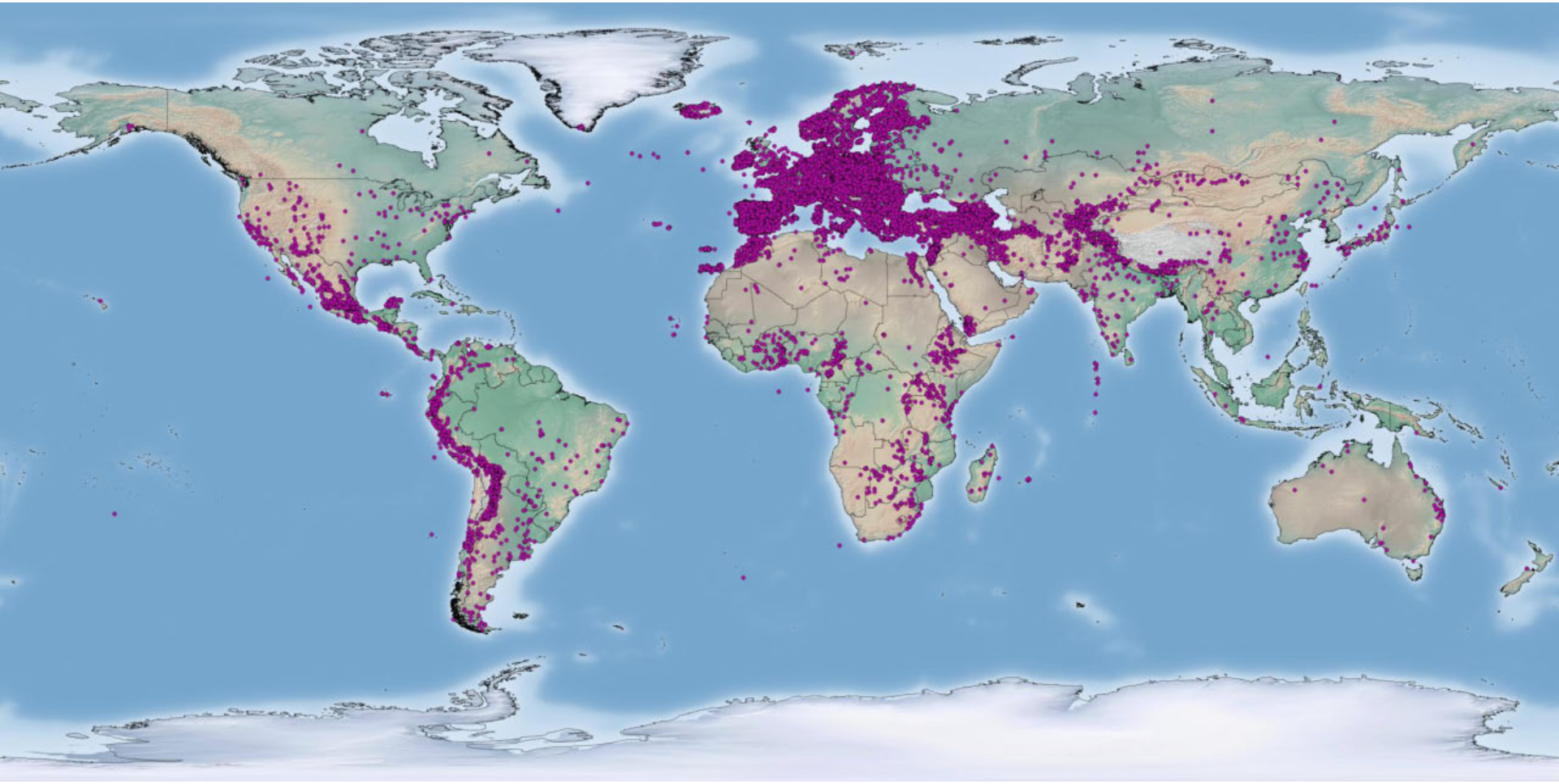
as of 2015-04-17

Countries of origin



as of 2015-04-17

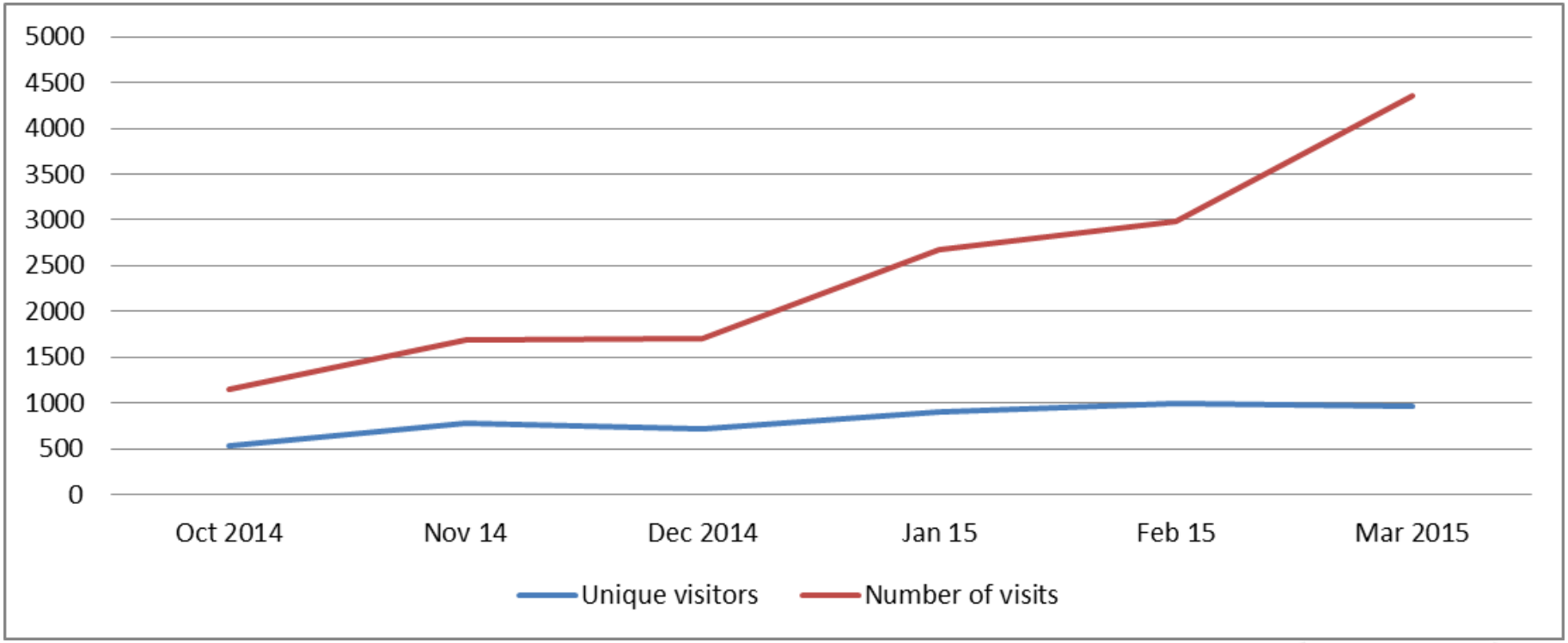
Regions of origin



as of 2015-04-17

Access by users

Month	Unique visitors	Number of visits	Pages
Oct 2014	541	1,143	17,808
Nov 14	783	1,692	63,315
Dec 2014	721	1,710	30,745
Jan 15	897	2,676	33,033
Feb 15	996	2,977	42,683
Mar 2015	960	4,354	38,658



TRANSFER TO NEW HOST

EURISCO transfer I

- October 2012:
  - Request for tenders for hosting EURISCO
- March 2013:
  - IPK won the bid
- May and October 2013
  - Preparatory meetings in Gatersleben and Rome
- Since 15th April 2014:
  - Sub-contract with Bioversity International
  - Transfer of the EURISCO responsibilities (04-09/2014)

EURISCO transfer II

- Preparatory meetings
  - Challenges:
    - Outdated systems
    - Insufficient technical documentation
    - High costs for transfer of as-is status
  - Decisions:

    →No transfer of existing solutions

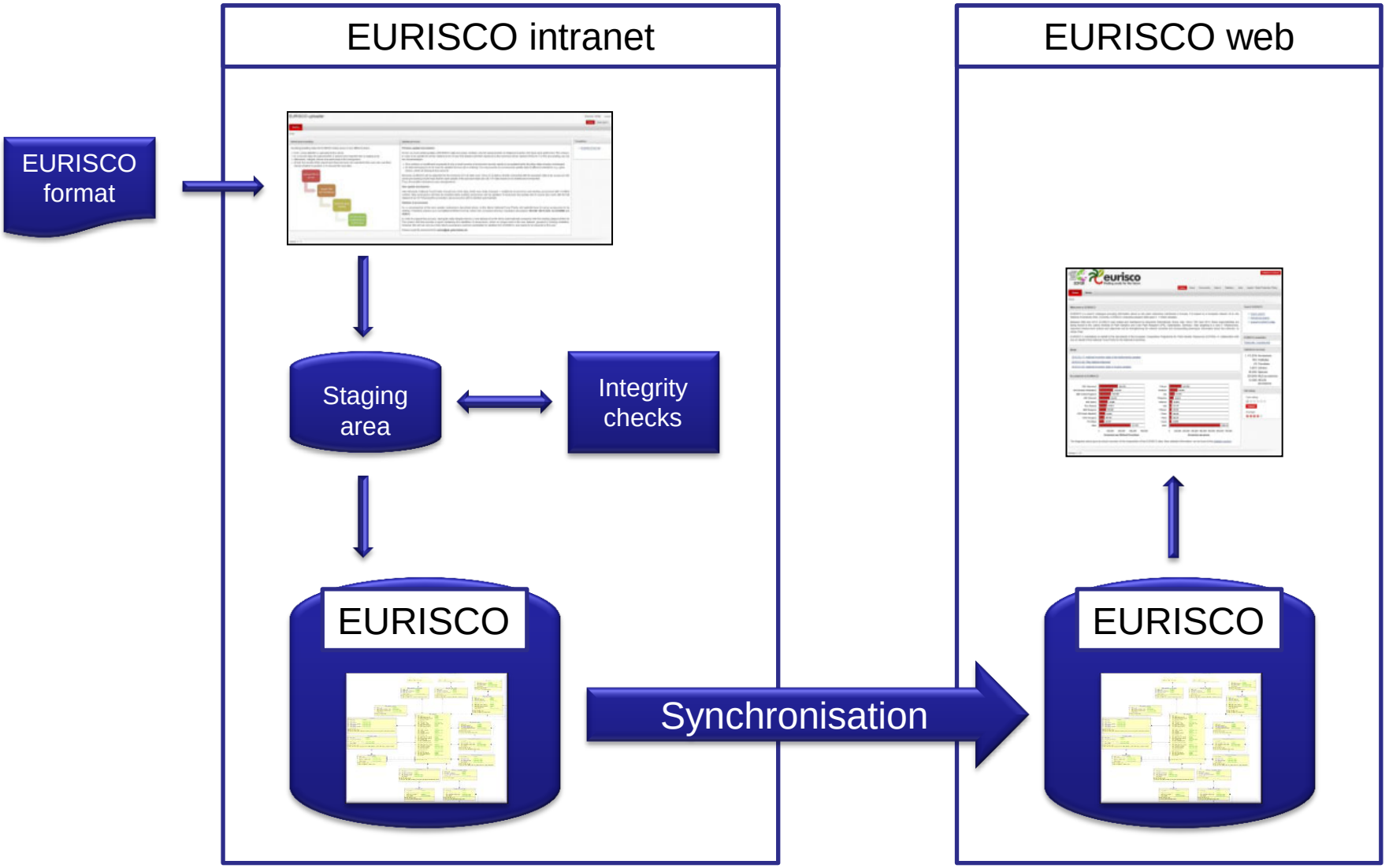
    →New development from scratch

EURISCO transfer III

- Analysis of the old web application
- Reengineering of database schema for web application
- Import (and cleansing) of current data set
  - Migration path MySQL → Oracle RDBMS

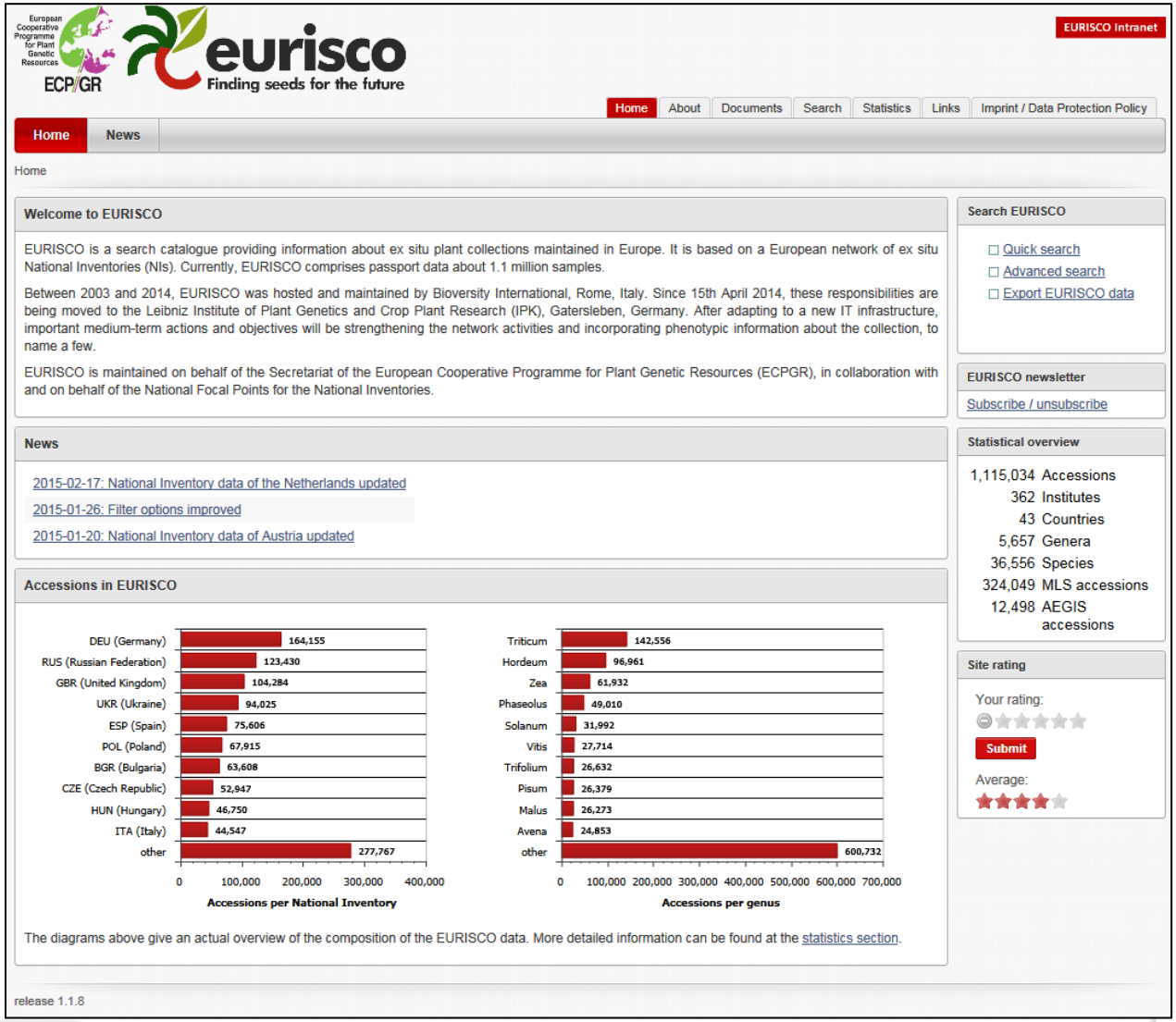
- New web application for searching EURISCO data
  - PL/SQL for functionality; APEX for rendering
  - First release with basic functionality published 09/2014
  - Will be improved continuously
- New web application for updating NI data
  - PL/SQL for functionality; APEX for rendering

New architecture

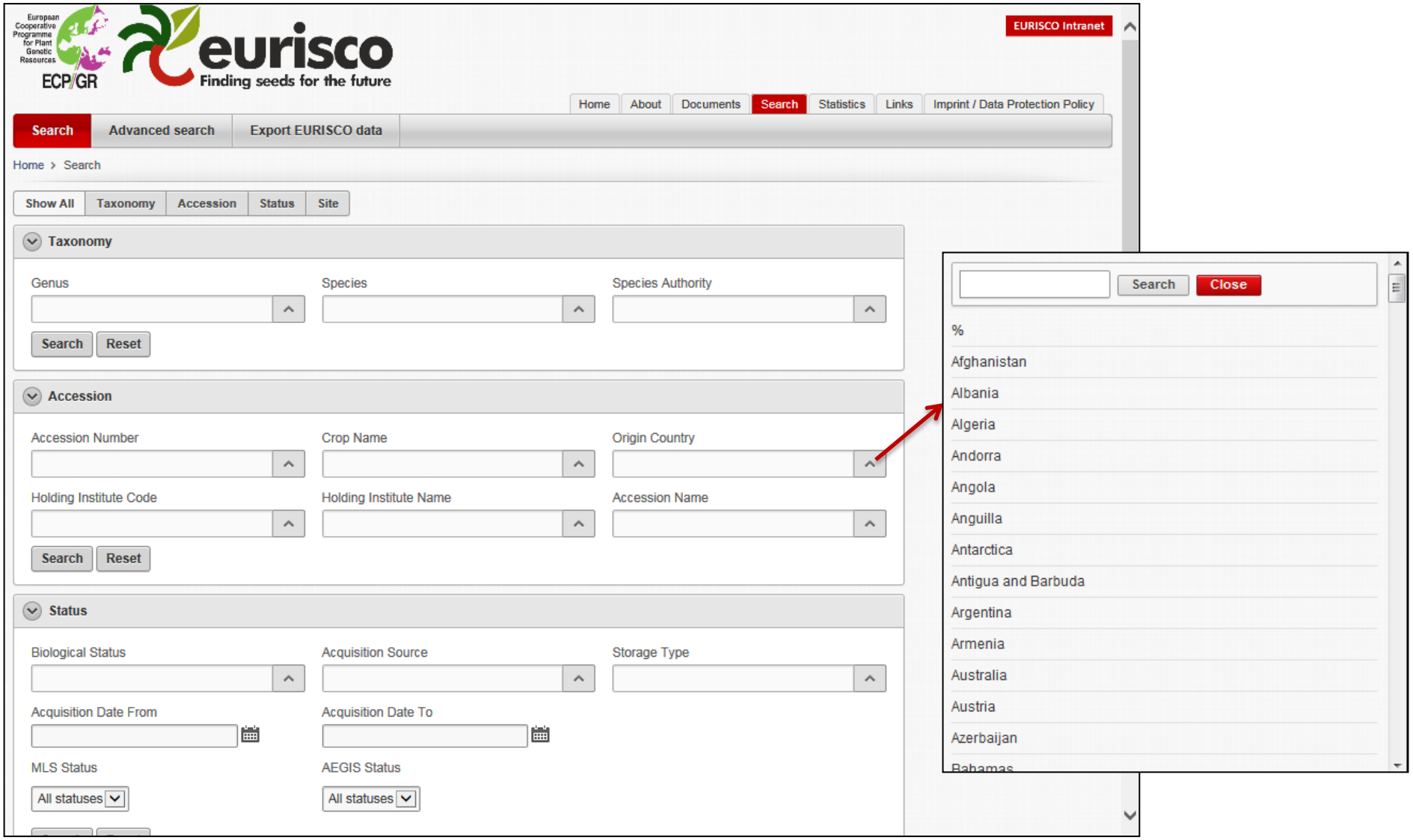


EURISCO WEB

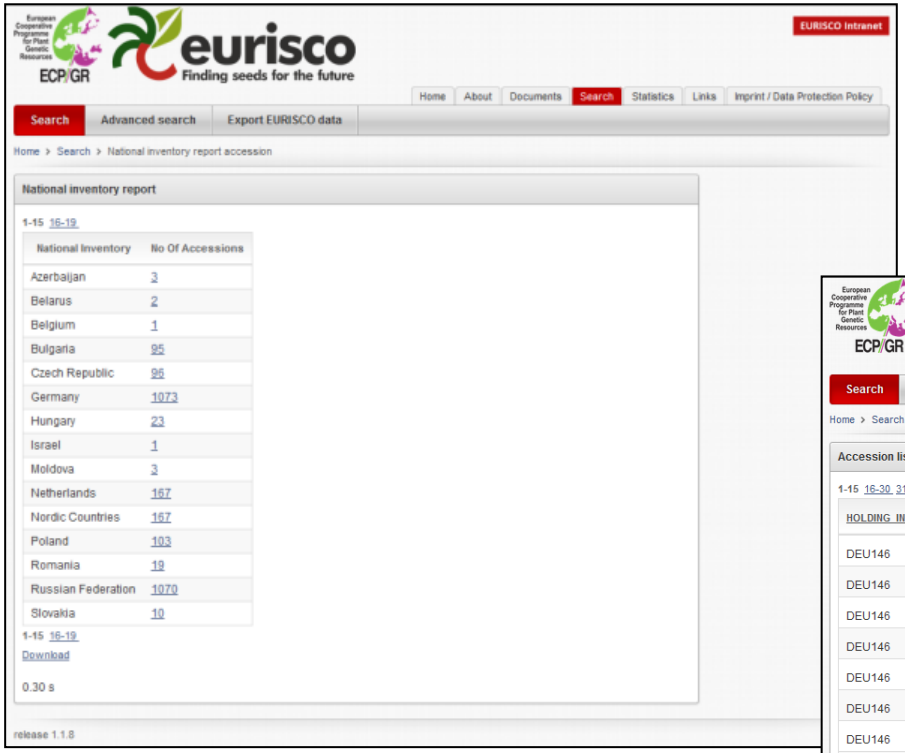
New web application

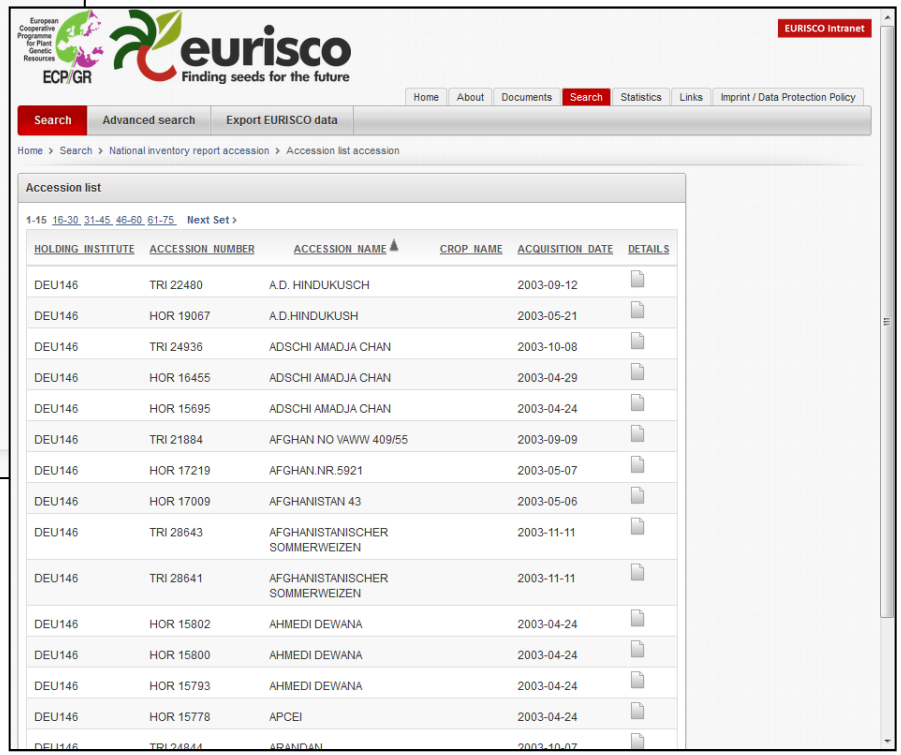


Search form

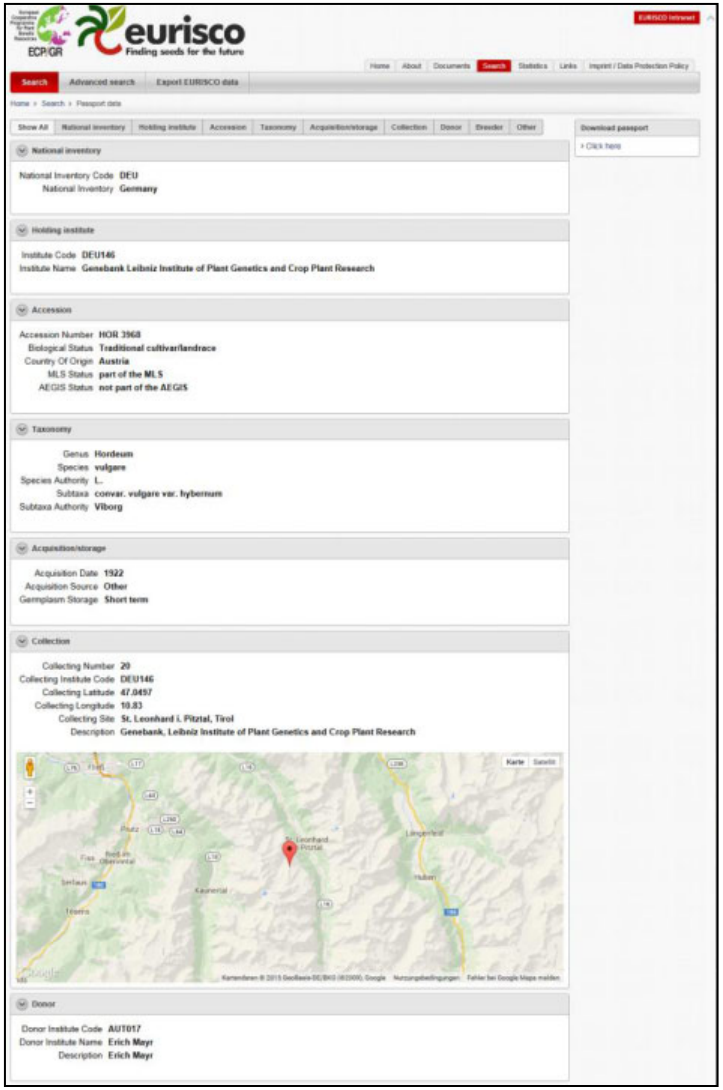

European Cooperative Programme for Plant Genetic Resources
ECP/GR
eurisco
Finding seeds for the future
EURISCO Intranet
Home
About
Documents
Search
Statistics
Links
Imprint / Data Protection Policy
Search
Advanced search
Export EURISCO data
Home › Search
Show All
Taxonomy
Accession
Status
Site
Taxonomy
Genus
Species
Species Authority
Search
Reset
Accession
Accession Number
Crop Name
Origin Country
Holding Institute Code
Holding Institute Name
Accession Name
Search
Reset
Status
Biological Status
Acquisition Source
Storage Type
Acquisition Date From
Acquisition Date To
MLS Status
AEGIS Status
All statuses
All statuses
Search
Close
%
Afghanistan
Albania
Algeria
Andorra
Angola
Anguilla
Antarctica
Antigua and Barbuda
Argentina
Armenia
Australia
Austria
Azerbaijan
Bahamas

Search results





Accession details



National inventory

National Inventory Code **DEU**
National Inventory **Germany**

Holding institute

Institute Code **DEU146**
Institute Name **Genebank Leibniz Institute of Plant Genetics and Crop Plant Research**

Accession

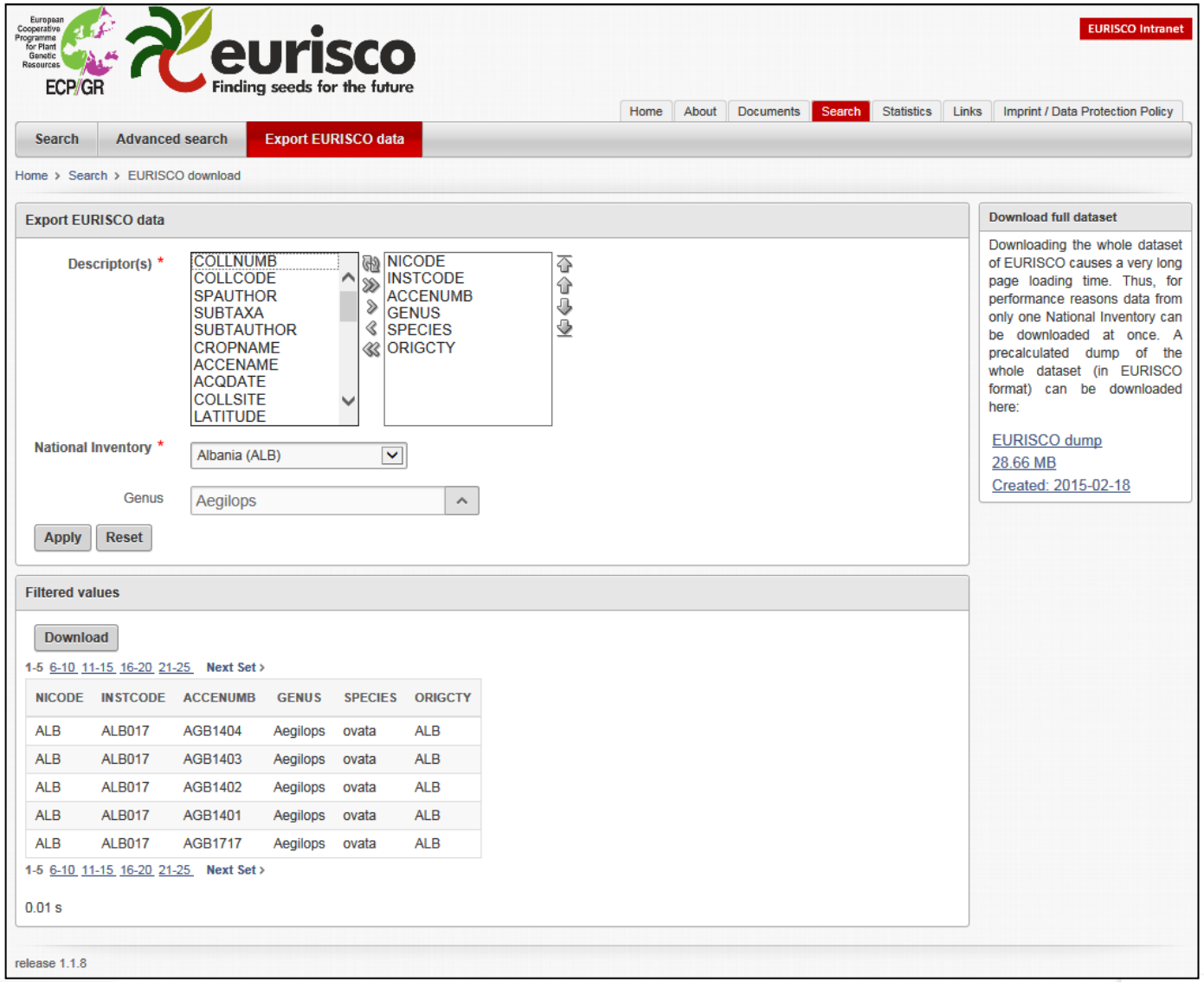
Accession Number **HOR 3968**
Biological Status **Traditional cultivar/landrace**
Country Of Origin **Austria**
MLS Status **part of the MLS**
AEGIS Status **not part of the AEGIS**

Taxonomy

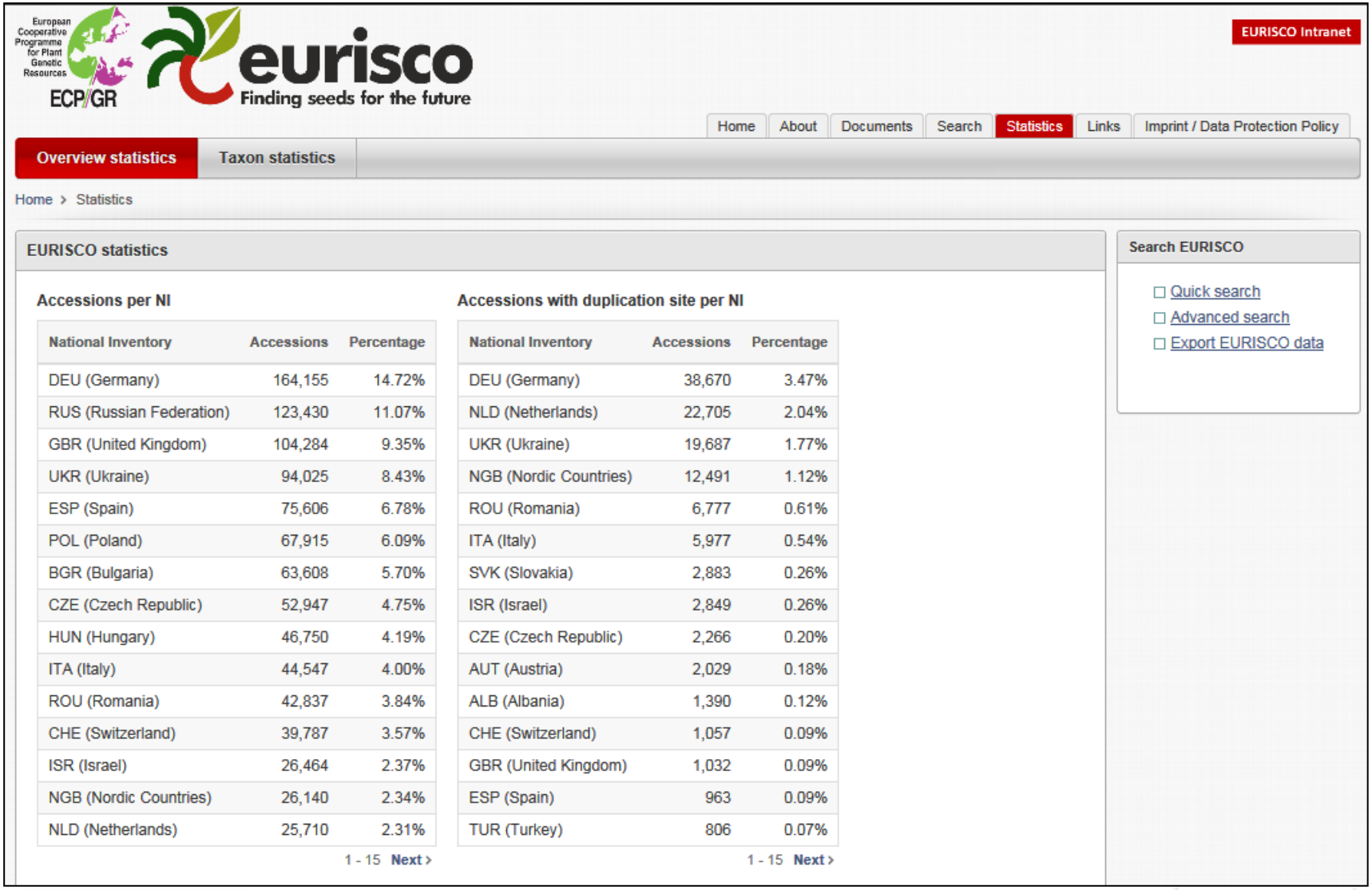
Genus **Hordeum**
Species **vulgare**
Species Authority **L.**
Subtaxa **convar. vulgare var. hybernum**
Subtaxa Authority **Viborg**

...

User specific export



Statistics



European Cooperative Programme for Plant Genetic Resources
ECP/GR
eurisco
Finding seeds for the future

EURISCO Intranet

Home | About | Documents | Search | Statistics | Links | Imprint / Data Protection Policy

Overview statistics | Taxon statistics

Home › Statistics

EURISCO statistics

Accessions per NI

National Inventory	Accessions	Percentage
DEU (Germany)	164,155	14.72%
RUS (Russian Federation)	123,430	11.07%
GBR (United Kingdom)	104,284	9.35%
UKR (Ukraine)	94,025	8.43%
ESP (Spain)	75,606	6.78%
POL (Poland)	67,915	6.09%
BGR (Bulgaria)	63,608	5.70%
CZE (Czech Republic)	52,947	4.75%
HUN (Hungary)	46,750	4.19%
ITA (Italy)	44,547	4.00%
ROU (Romania)	42,837	3.84%
CHE (Switzerland)	39,787	3.57%
ISR (Israel)	26,464	2.37%
NGB (Nordic Countries)	26,140	2.34%
NLD (Netherlands)	25,710	2.31%

1 - 15 Next ›

Accessions with duplication site per NI

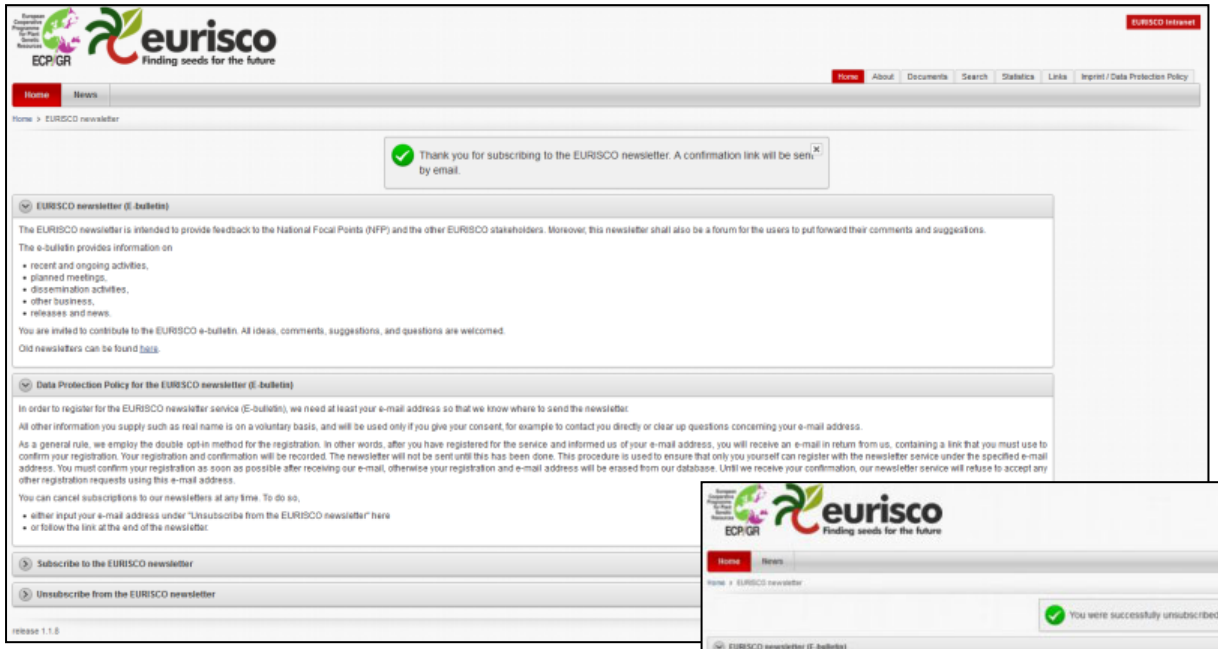
National Inventory	Accessions	Percentage
DEU (Germany)	38,670	3.47%
NLD (Netherlands)	22,705	2.04%
UKR (Ukraine)	19,687	1.77%
NGB (Nordic Countries)	12,491	1.12%
ROU (Romania)	6,777	0.61%
ITA (Italy)	5,977	0.54%
SVK (Slovakia)	2,883	0.26%
ISR (Israel)	2,849	0.26%
CZE (Czech Republic)	2,266	0.20%
AUT (Austria)	2,029	0.18%
ALB (Albania)	1,390	0.12%
CHE (Switzerland)	1,057	0.09%
GBR (United Kingdom)	1,032	0.09%
ESP (Spain)	963	0.09%
TUR (Turkey)	806	0.07%

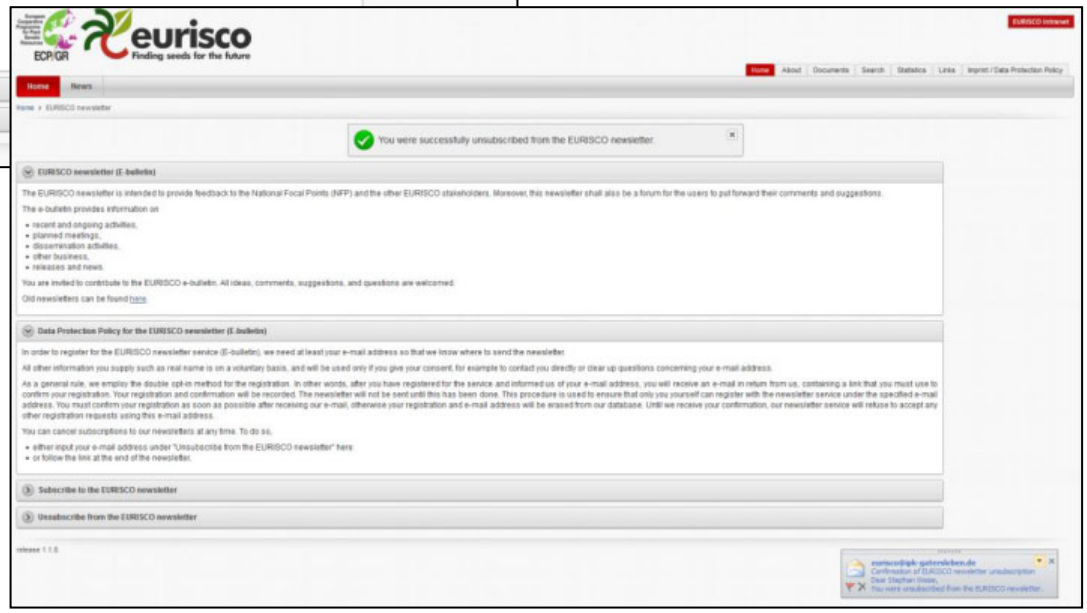
1 - 15 Next ›

Search EURISCO

- Quick search
- Advanced search
- Export EURISCO data

Newsletter subscription





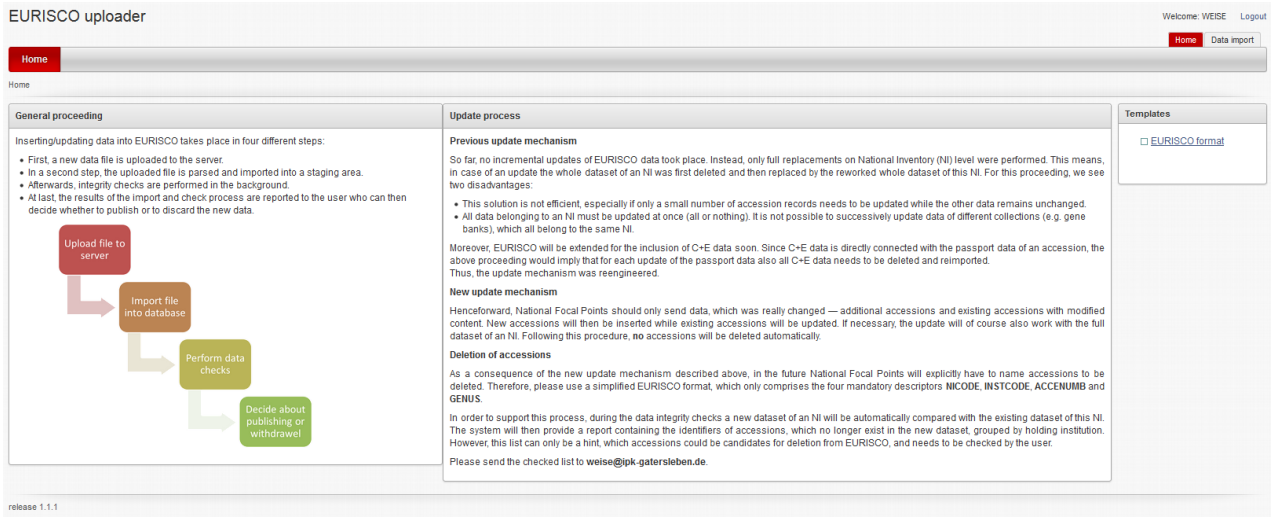
EURISCO web application

- Same DB schema for EURISCO data
- For the web application
  - Additional tables
  - Materialised views
  - PL/SQL procedures
  - JavaScript functions
- Features
  - Search (simple/advanced)
  - User specific export + full dump
  - Statistics
  - Document archive
  - Newsletter with double opt-in
  - Other information + links

EURISCO INTRANET

EURISCO intranet I

- Development of new import component for NIs
  - Web interface with Oracle APEX
  - Seven PL/SQL packages with 77 procedures for uploading, checking and updating data
  - Implementation of incremental updates



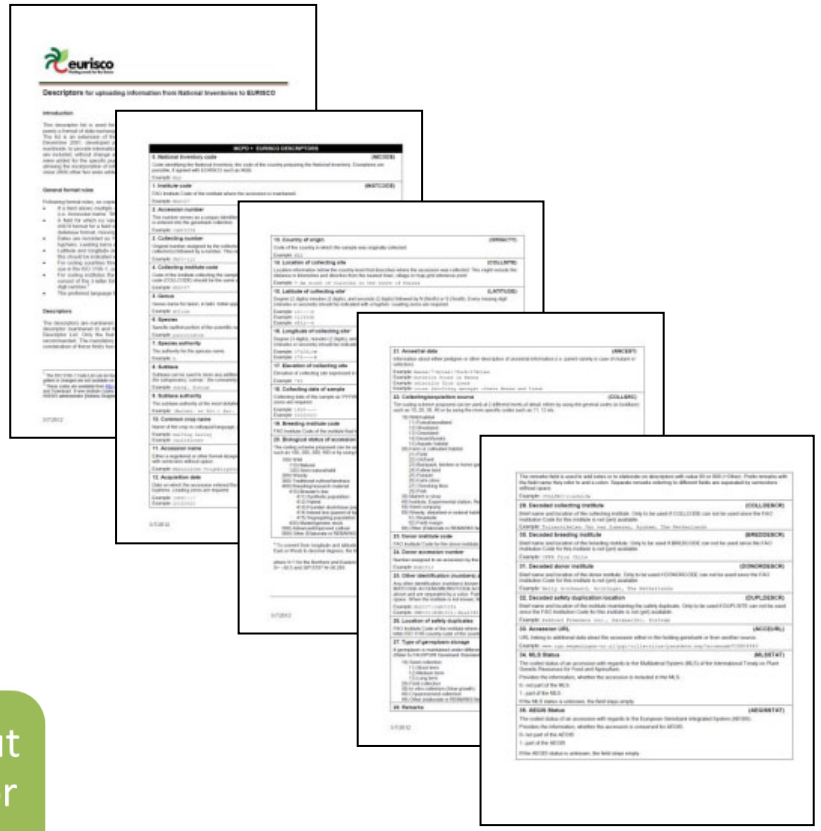
EURISCO intranet II

Upload file to server

Import file into database

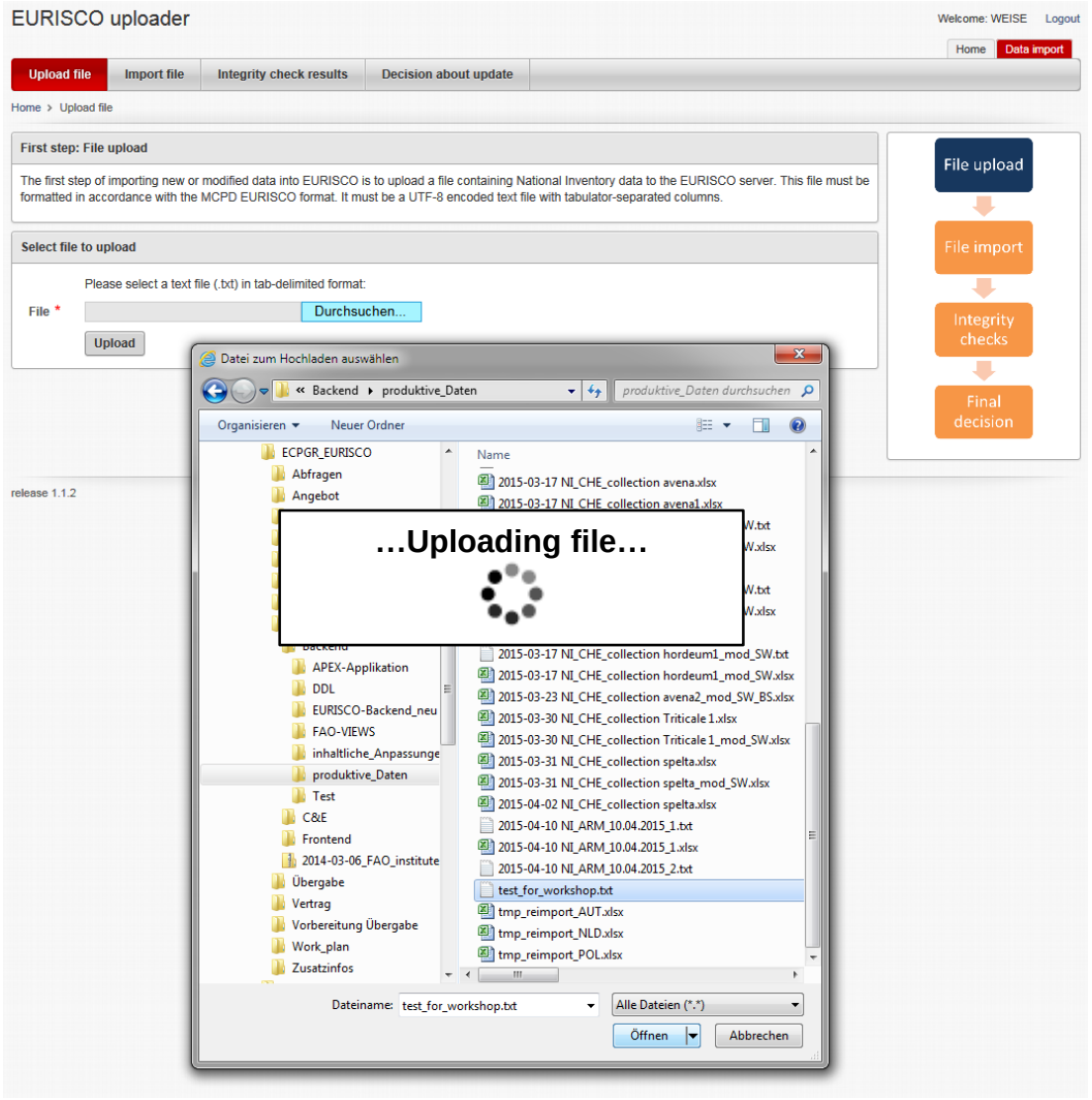
Perform data checks

Decide about publishing or withdrawel

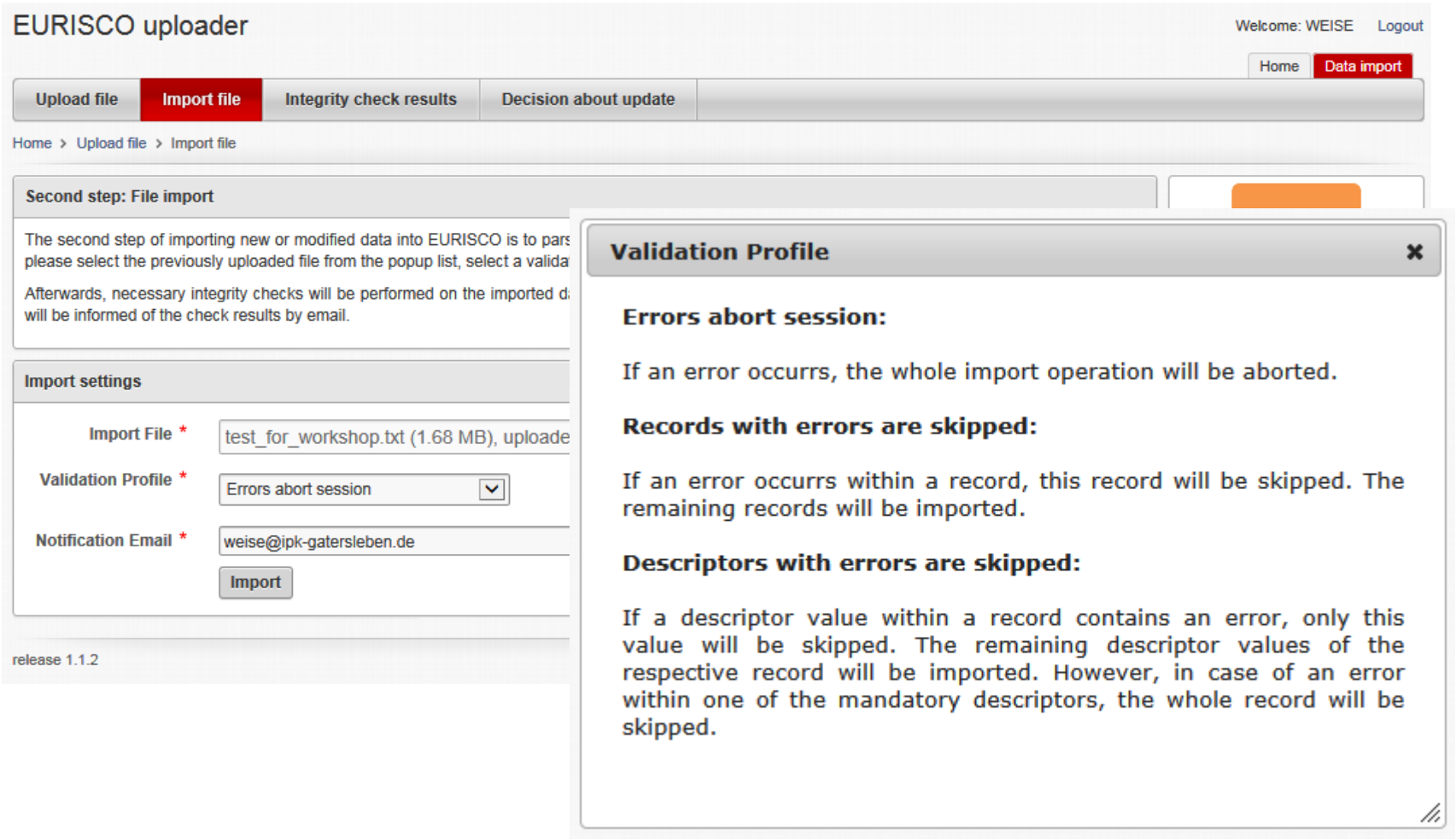


File upload

- Text file
- Tab separated
- UTF-8



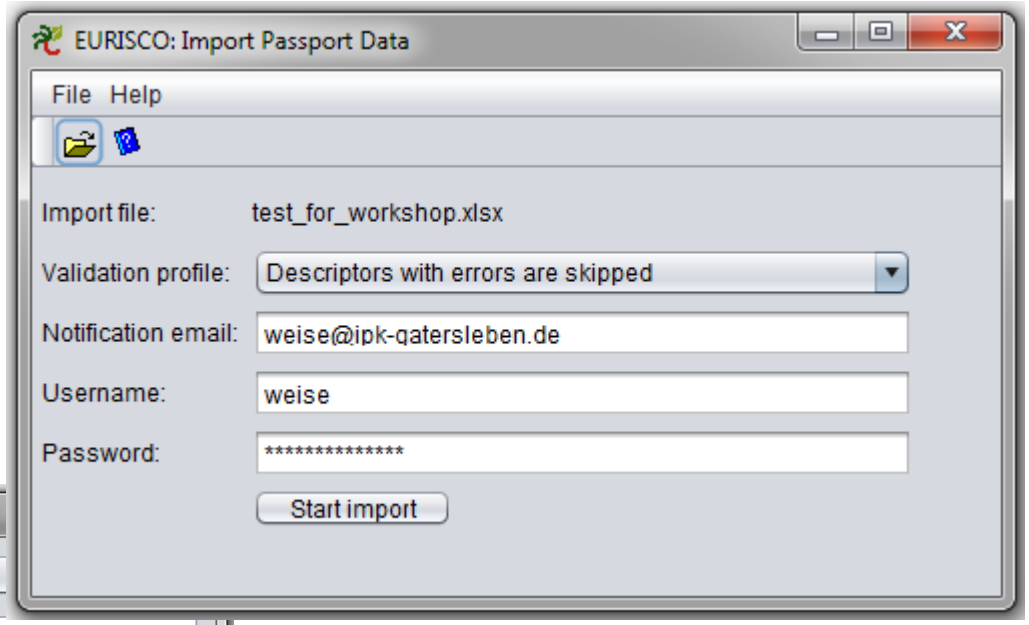
File import

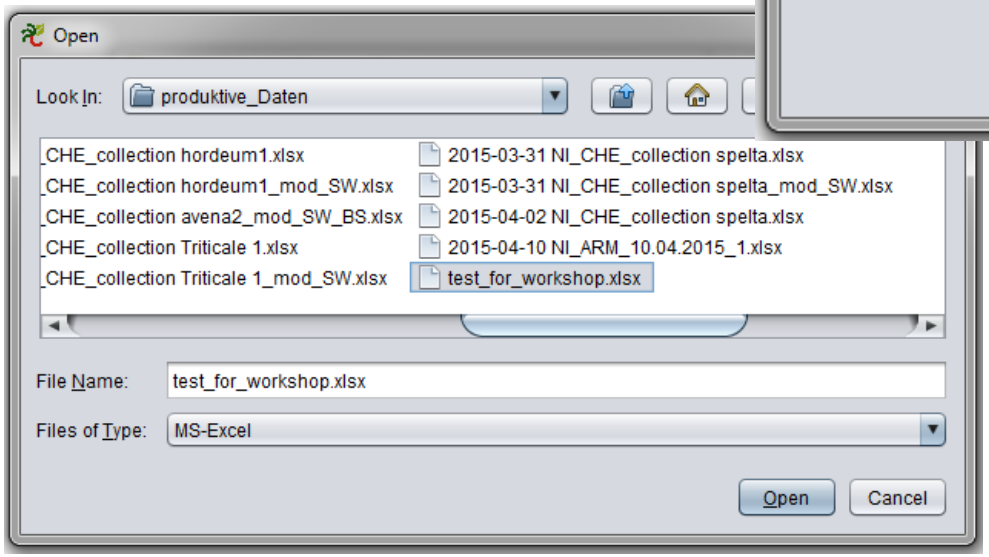

EURISCO uploader
Welcome: WEISE
Logout
Home
Data import
Upload file
Import file
Integrity check results
Decision about update
Home › Upload file › Import file
Second step: File import
The second step of importing new or modified data into EURISCO is to pars
please select the previously uploaded file from the popup list, select a valida
Afterwards, necessary integrity checks will be performed on the imported d
will be informed of the check results by email.
Import settings
Import File *
test_for_workshop.txt (1.68 MB), uploade
Validation Profile *
Errors abort session
Notification Email *
weise@ipk-gatersleben.de
Import
release 1.1.2
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.

Java-based upload I

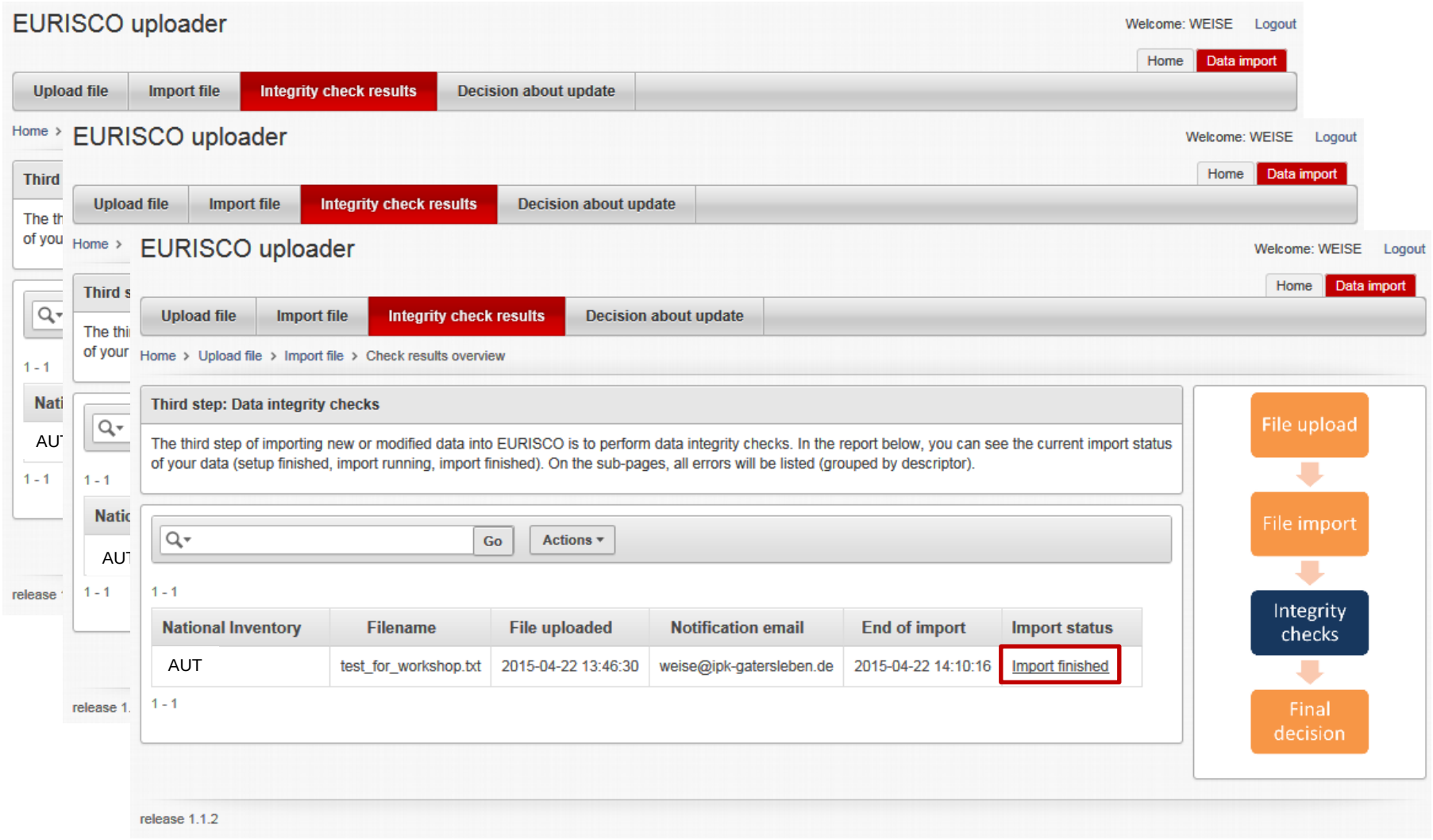
- So far: Web-based upload
  - Upload via tab-separated text file, UTF-8 encoded
  - Often problems with columns separators + character encoding
- That‘s why: Java-based Excel upload
  - Summarises file upload + data import
  - User gets informed when the integrity checks are finished

Java-based upload II

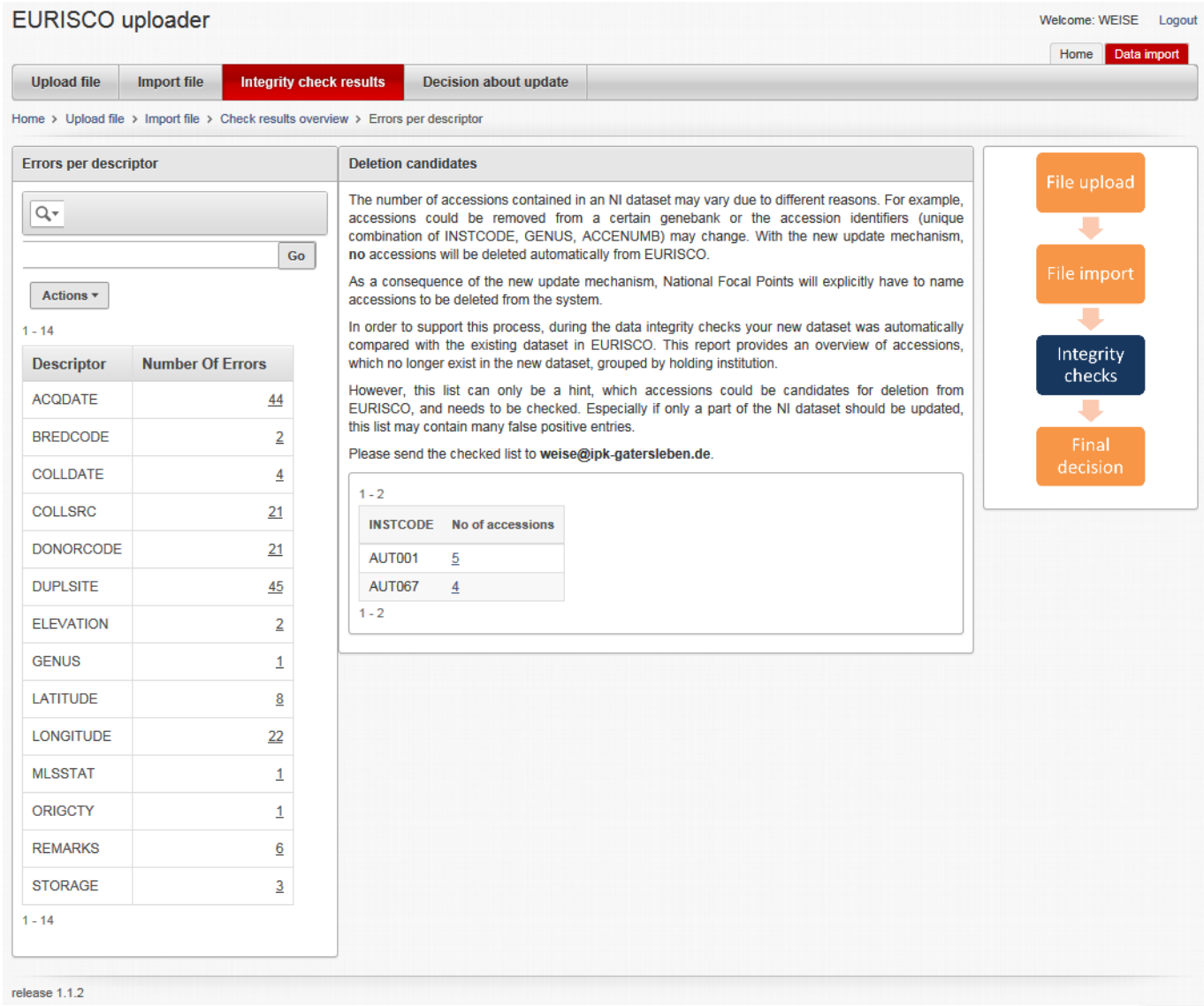




Integrity checks I

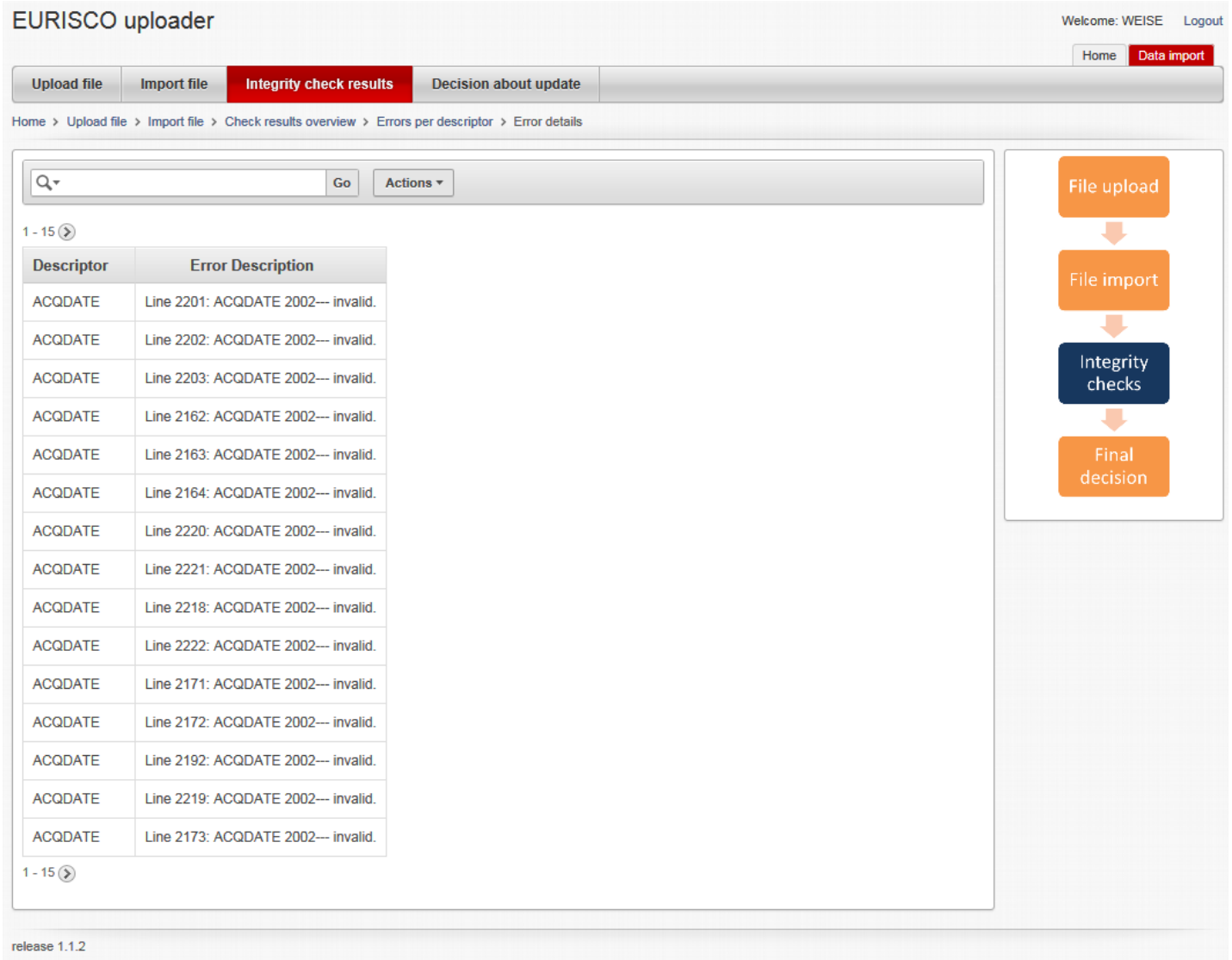


Integrity checks II – error report



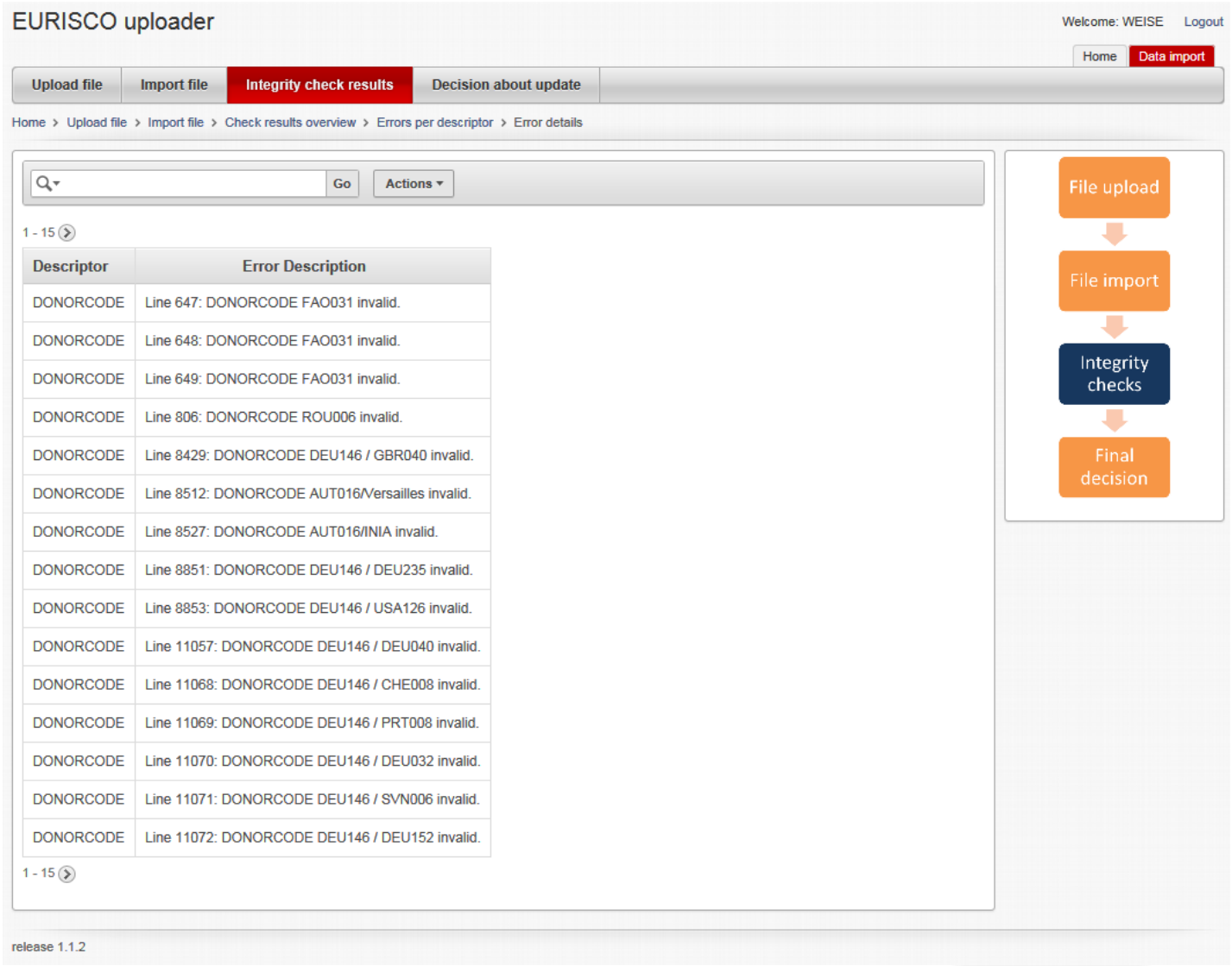
Integrity checks III - example

- Wrong date format



Integrity checks IV - example

- Invalid or multiple donor codes

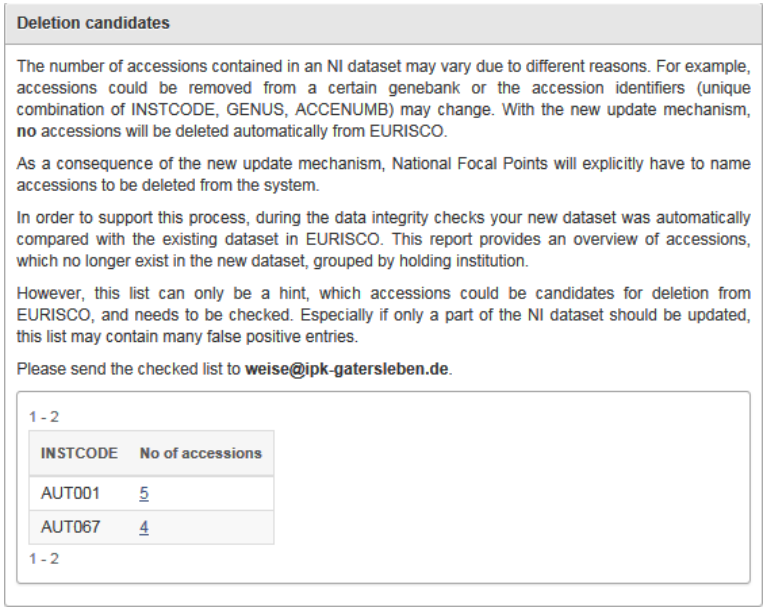


Incremental updates I

- 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.

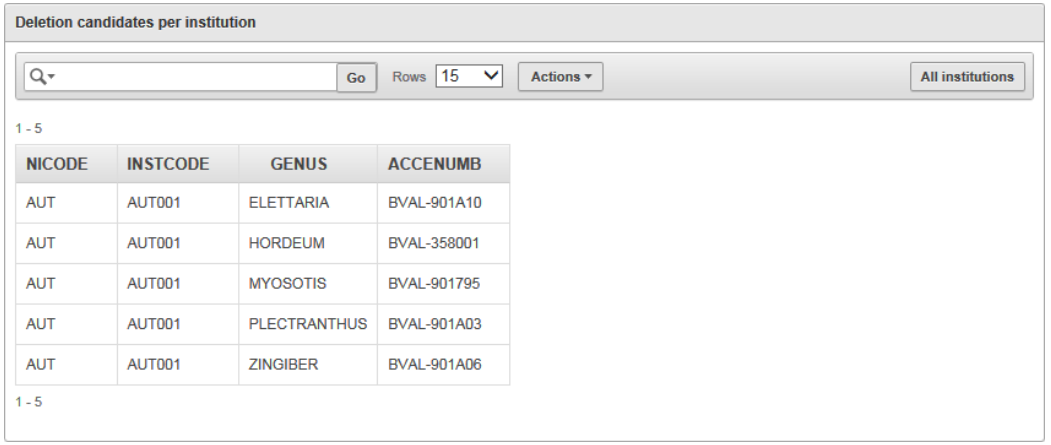
However, this list can only be a hint, which accessions could be candidates for deletion from EURISCO, and needs to be checked. Especially if only a part of the NI dataset should be updated, this list may contain many false positive entries.

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

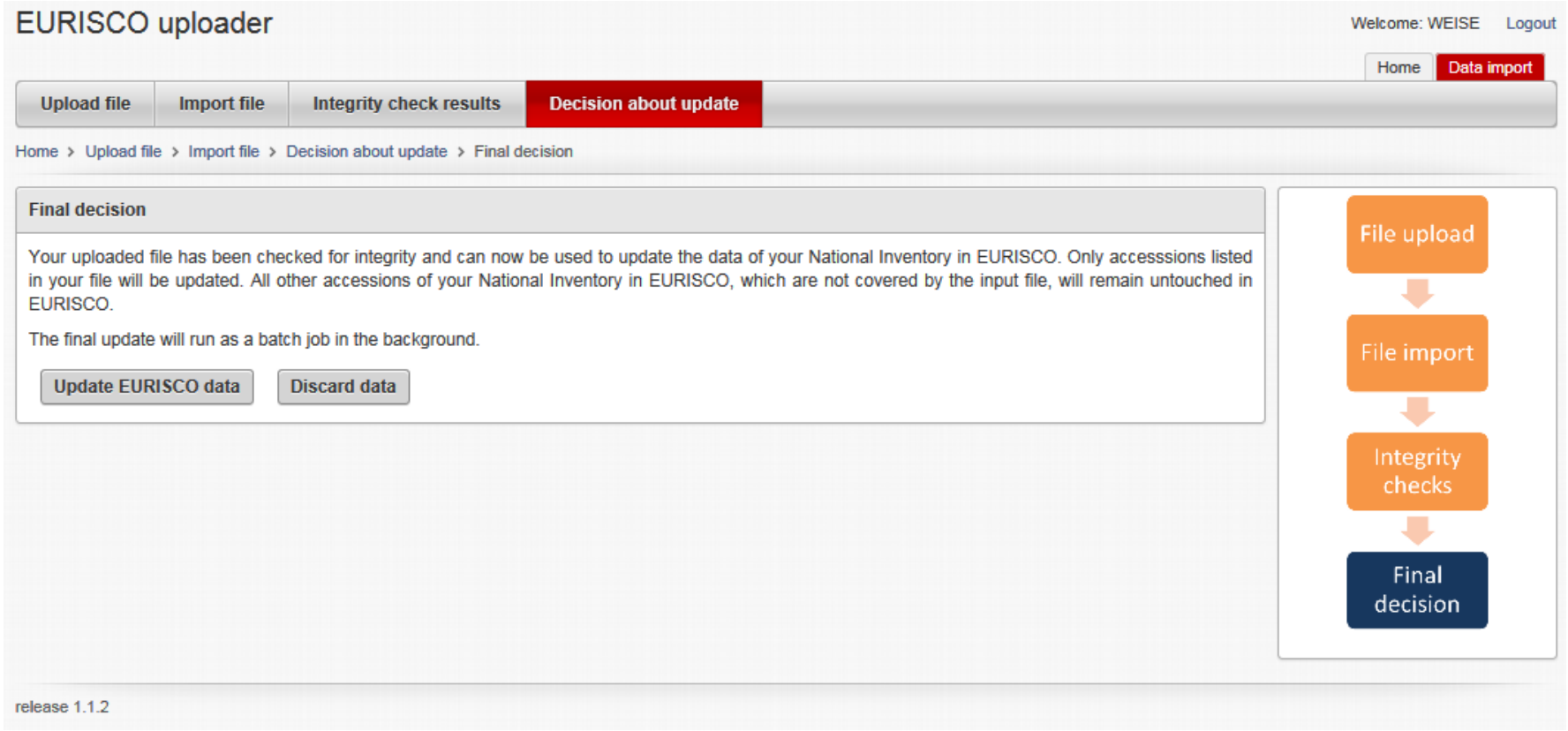
Go | Rows 15 | Actions | All institutions

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



FUTURE OF EURISCO

Increase data quantity/quality I

- Data quantity
  - Inclusion of additional passport data
    - ~250,000 accessions (van Hintum, 2014)
- Data quality
  - Increase frequency of updates
    - Avg. age 1.16 years
    - Oldest 10% avg. 4.96 years

    (van Hintum, 2014)

The documentation of Plant Genetic Resources in Europe

Theo van Hintum, Centre for Genetic Resources, The Netherlands (CGN)

April, 2014

This is a personal view as input for the discussions at the workshop of the Documentation and Information Working Group 'Tailoring the Documentation of Plant Genetic Resources in Europe to the Needs of the User' to be held 20-22 May '14 in Prague, Czech Republic. It provides a conceptual background to the issues to be discussed and formulates a number of draft resolution.

Focus

The current *ex situ* PGR documentation landscape consists of very many 'data sources', i.e., documentation systems of germplasm collections. The passport information from these systems is to a large extent collected by National Focal Points (NFPs) to create National Inventories (NIs). The information in NIs is expected to be regularly uploaded to EURISCO that thus should always provide an overview of the content of these NIs and of the genetic resources in Europe. Parallel to this data-flow and depository, Central Crop Data Bases (CCDBs) have been created since the early days of the ECPGR Crop Working Groups, to collect passport and sometimes additional data on a crop specific basis, however many of these databases do not appear to be up to date or to provide information and features that are not already present in EURISCO.

On a global level, the GeneSys initiative tries to create an entry point to data on all PGR maintained in the world. Data providers and database managers operate in an environment with rapidly evolving technologies and policies. It is therefore expected that developments in information technology, sequencing technology and policies on access and benefit sharing (ABS) will have a large impact on PGR documentation.

The combined data from EURISCO and 46 accessible Central Crop Data Bases originate from 506 data sources in 43 countries. The largest data source, according to EURISCO is IPK in Germany with 128k accessions, followed by the Vavilov Institute in Russia with 123k accessions[1]. The number of accessions currently documented in EURISCO is 1065766, and the total number of accessions in Europe is expected to be around 1.3 million.

In this document some important issues related to the current situation and developments regarding the documentation of PGR in Europe will be discussed, and resolutions will be formulated. These resolutions are aimed at either the EURISCO Management (currently, since beginning of 2014 at IPK) that coordinates the network of NFPs and runs the EURISCO database and web interface, or the ECPGR Doc/Info Working Group, that acts as the steering committee for EURISCO and oversees the documentation activities of ECPGR.

The issues that are to be discussed include: quality and coverage of the passport data in EURISCO, characterisation and evaluation data in EURISCO, the future of CCDBs in relation to EURISCO, PGR Portals, the relation EURISCO – GeneSys, the relation

1

Increase data quantity/quality II

- Data quality
  - Improve taxonomic backbone of EURISCO
    - Management of taxon synonyms
    - Improvement of checks during import
      - GRIN, Catalogue of Life webservice

http://www.ars-grin.gov

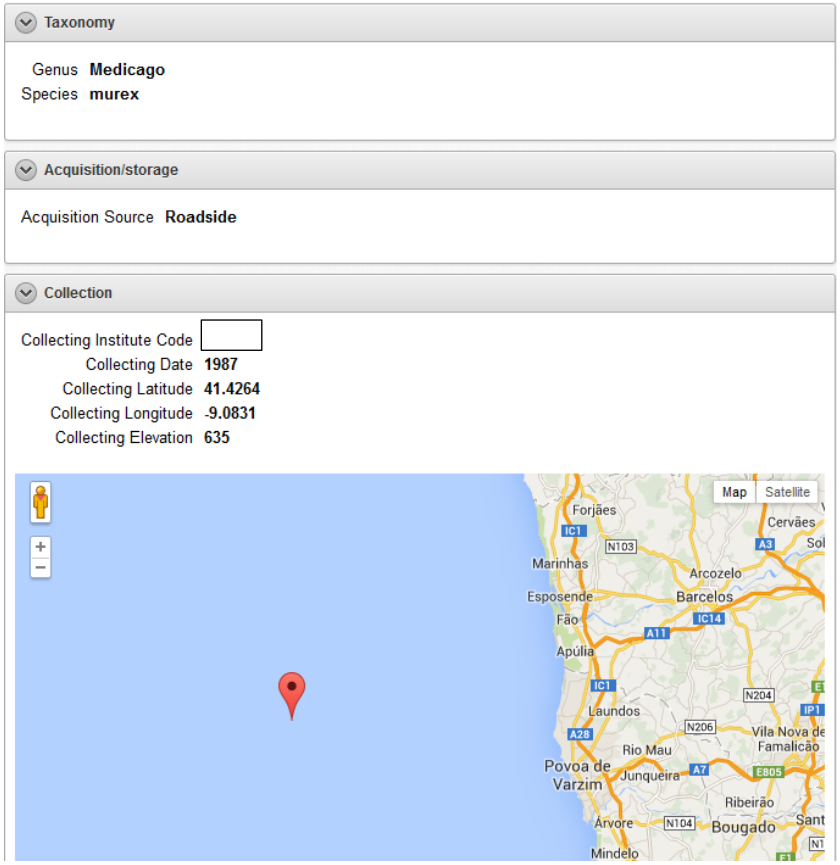
Catalogue of Life

http://www.catalogueoflife.org

  - Increase completeness of information
    - Often limited information about certain accessions

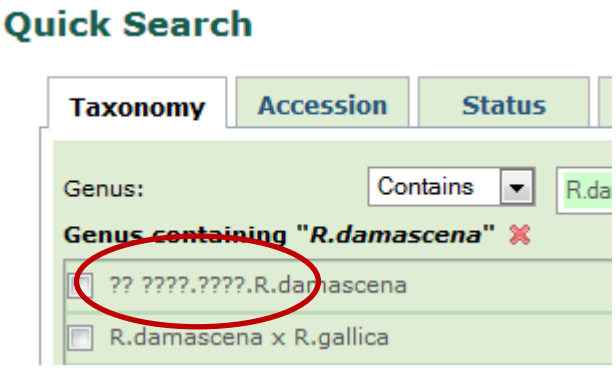
Increase data quantity/quality III

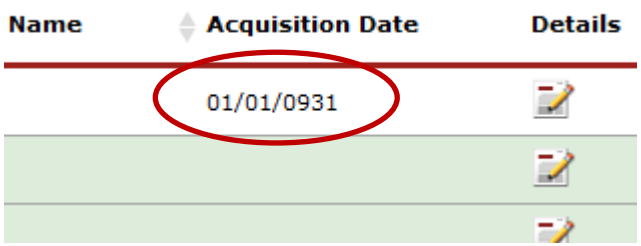
- Data quality
  - Improve location data quality

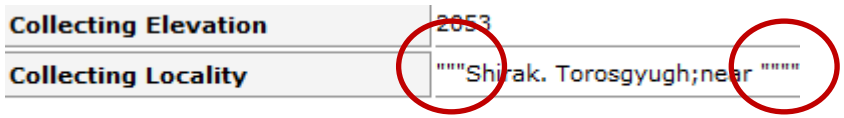


Increase data quantity/quality IV

- Data quality
  - Reduce inconsistencies

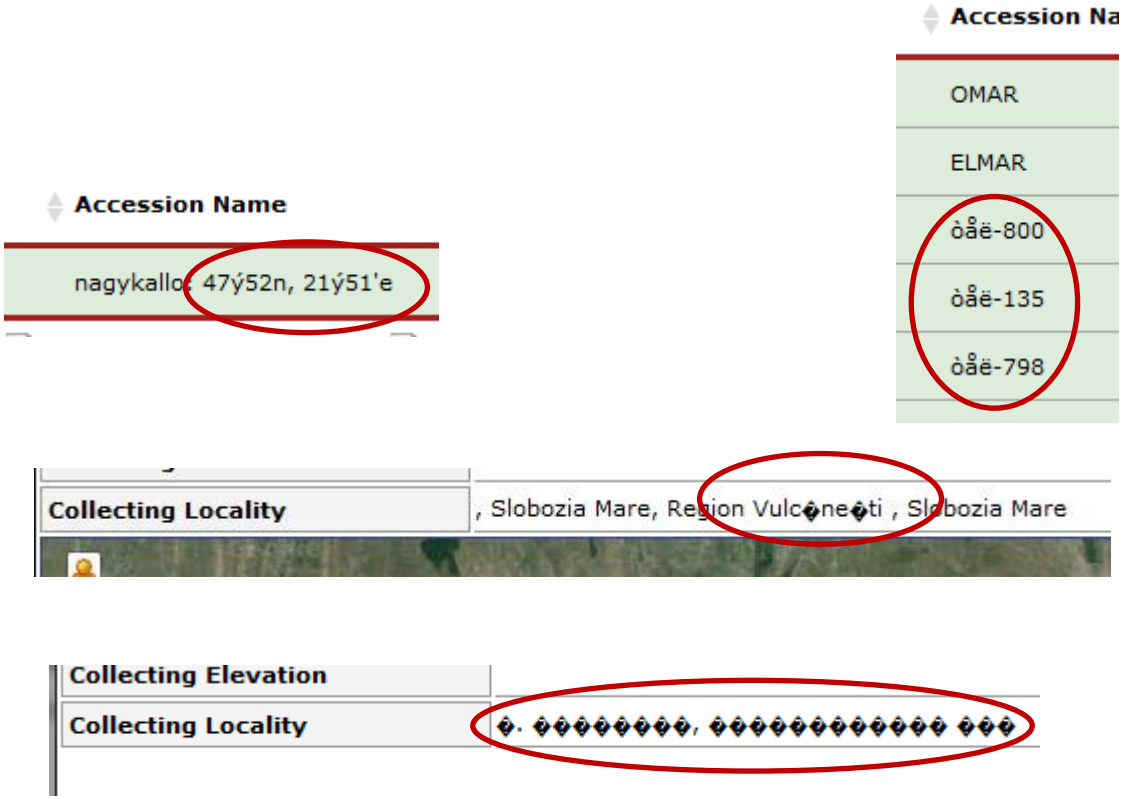






Increase data quantity/quality V

- Data quality
  - Fix encoding issues



Increase functionality

- Improve import mechanism
  - Nested user accounts
    - Data upload by holding institutes
    - Check and approval by NFPs
- Improve public web interface
  - Additional reports and download facilities (e.g. download by crop or genus)
  - Extend filter possibilities by additional fields
  - Enable case insensitive search for taxa
  - Provide a “free keyword search” over all alphanumeric fields

Aditional means of access

- Web services
  - Additional means of access
    - Data exchange with Genesys, GBIF etc.
  - Improvement of upload mechanism
    - Selective updates of certain accessions
- Mobile version

Extend for *in situ* data

- Will be documented in EURISCO
- Sufficient specification needed
- Data exchange formats need to be agreed

Extend for C&E data

- *Presentation on day 3*

DISSEMINATION

Dissemination

- Presentations
  - Doc&Info meeting, Prague, 2014
  - SEEDNet workshop, Ljubljana, 2014
- Short article submitted
- Regularly short information in ECPGR bulletin
- In preparation
  - Newsletter
  - Update of fact sheets
  - EURISCO poster/flyer on conferences

THANK YOU FOR YOUR ATTENTION