

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

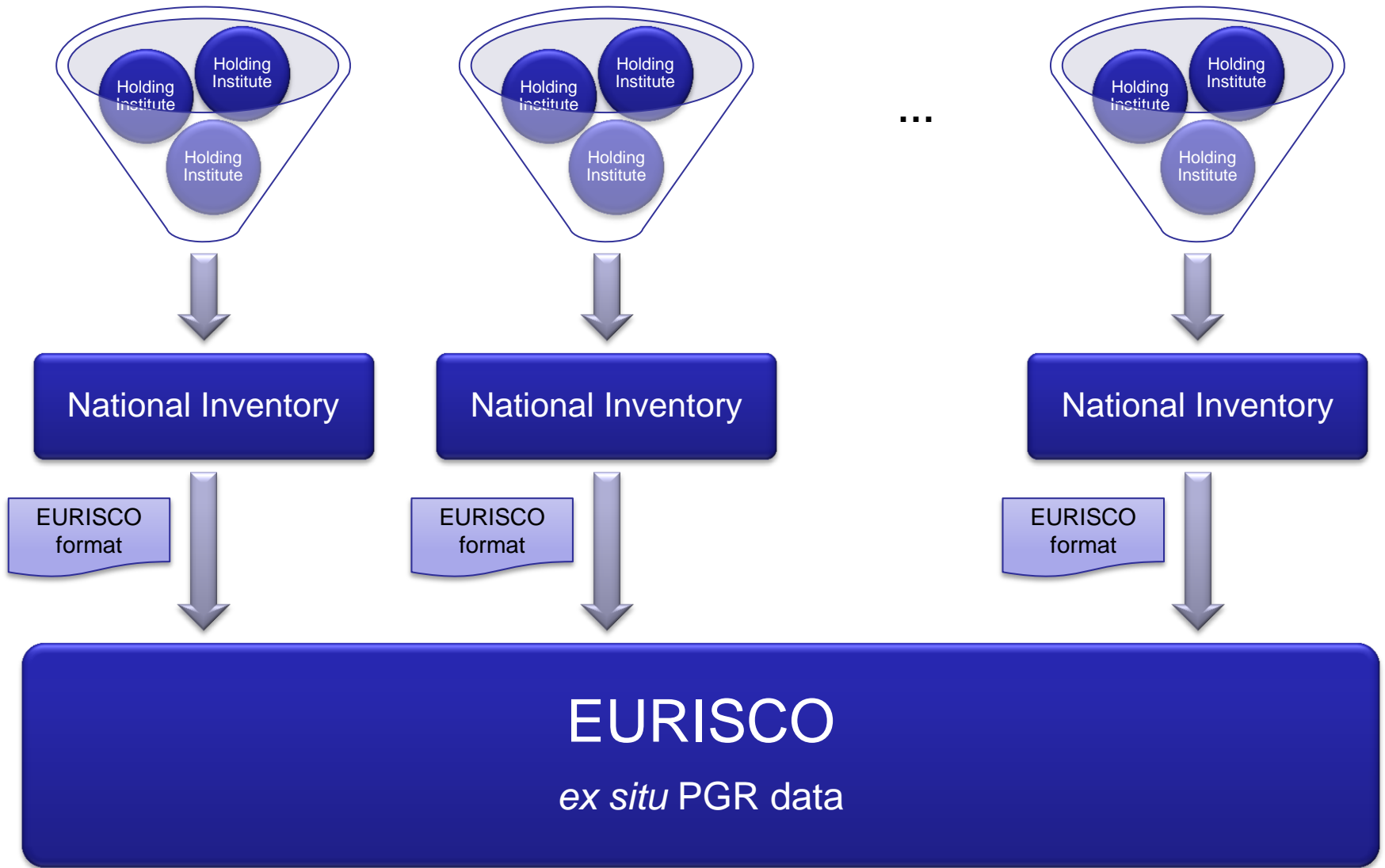
Introduction II

- 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

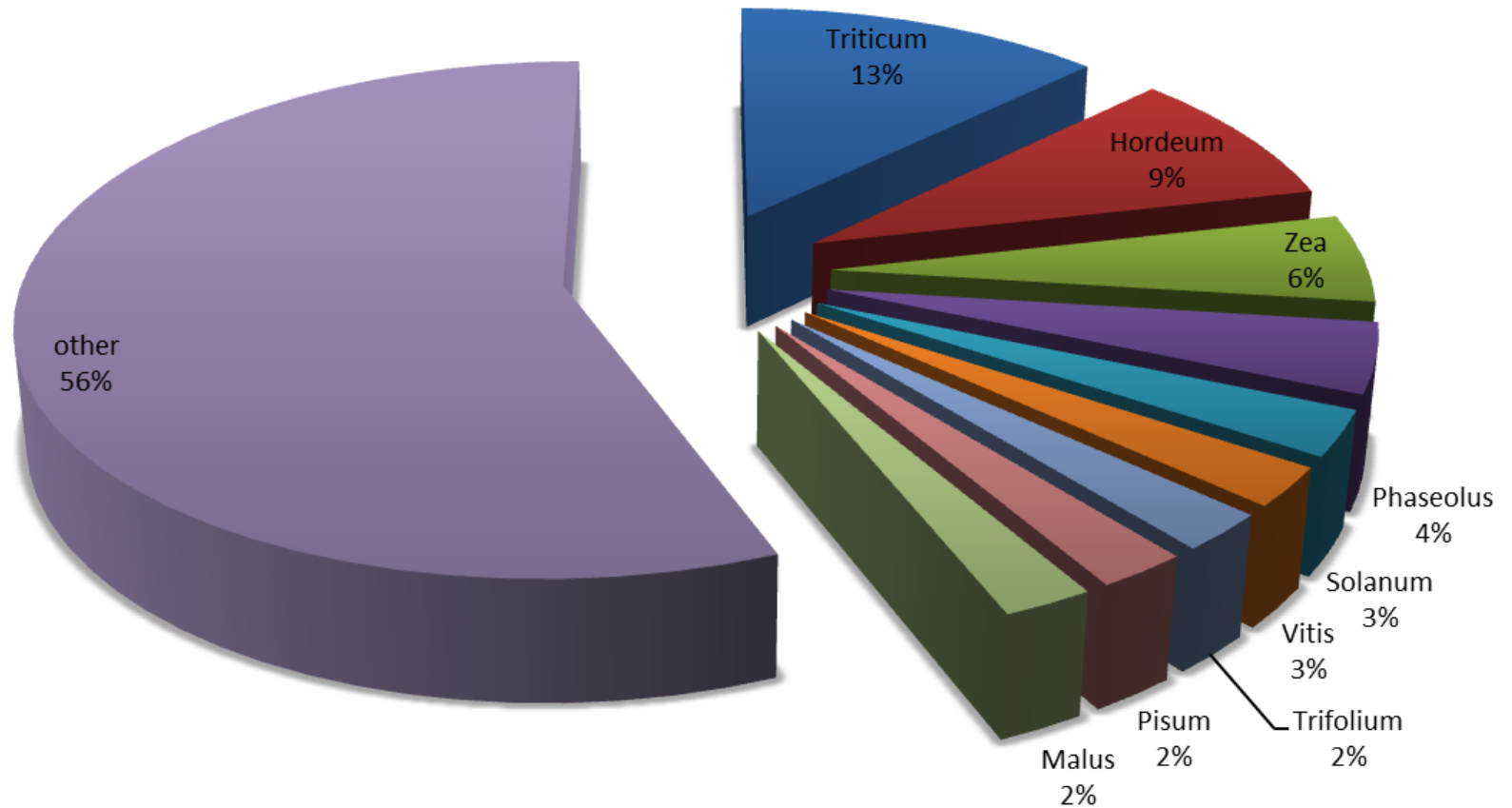
as of 2015-04-17

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

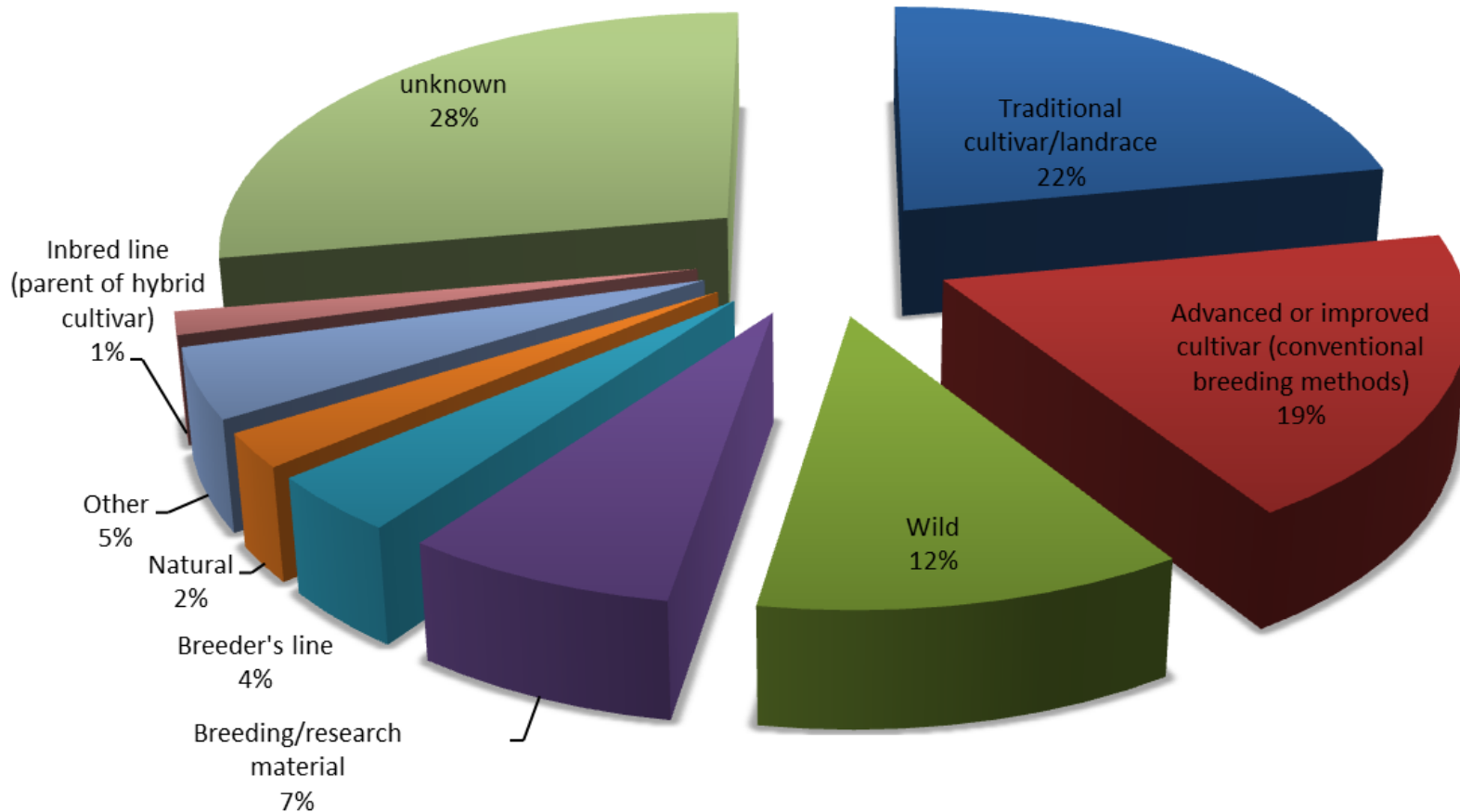
as of 2015-04-17

Genera



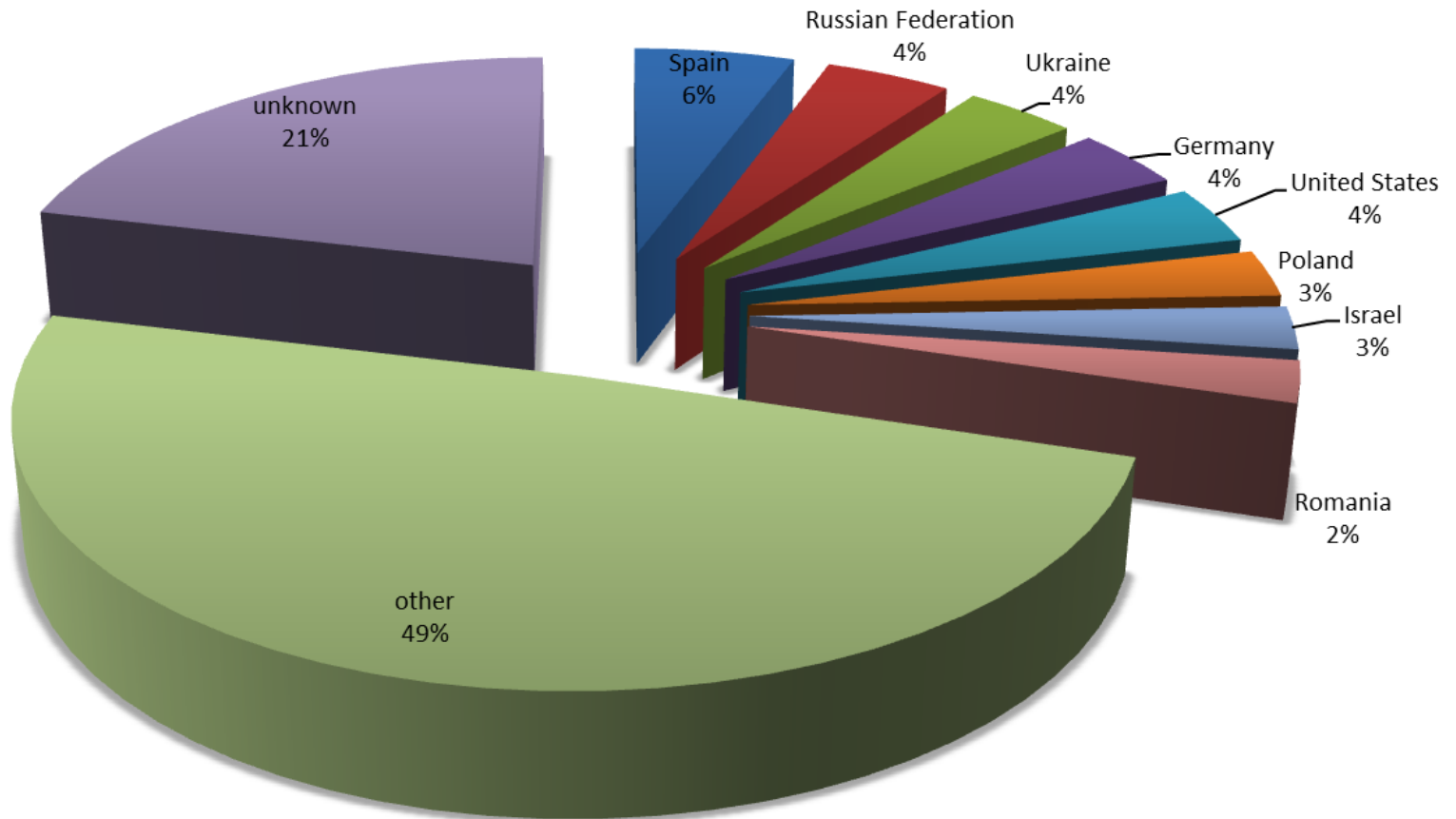
as of 2015-04-17

Biological status



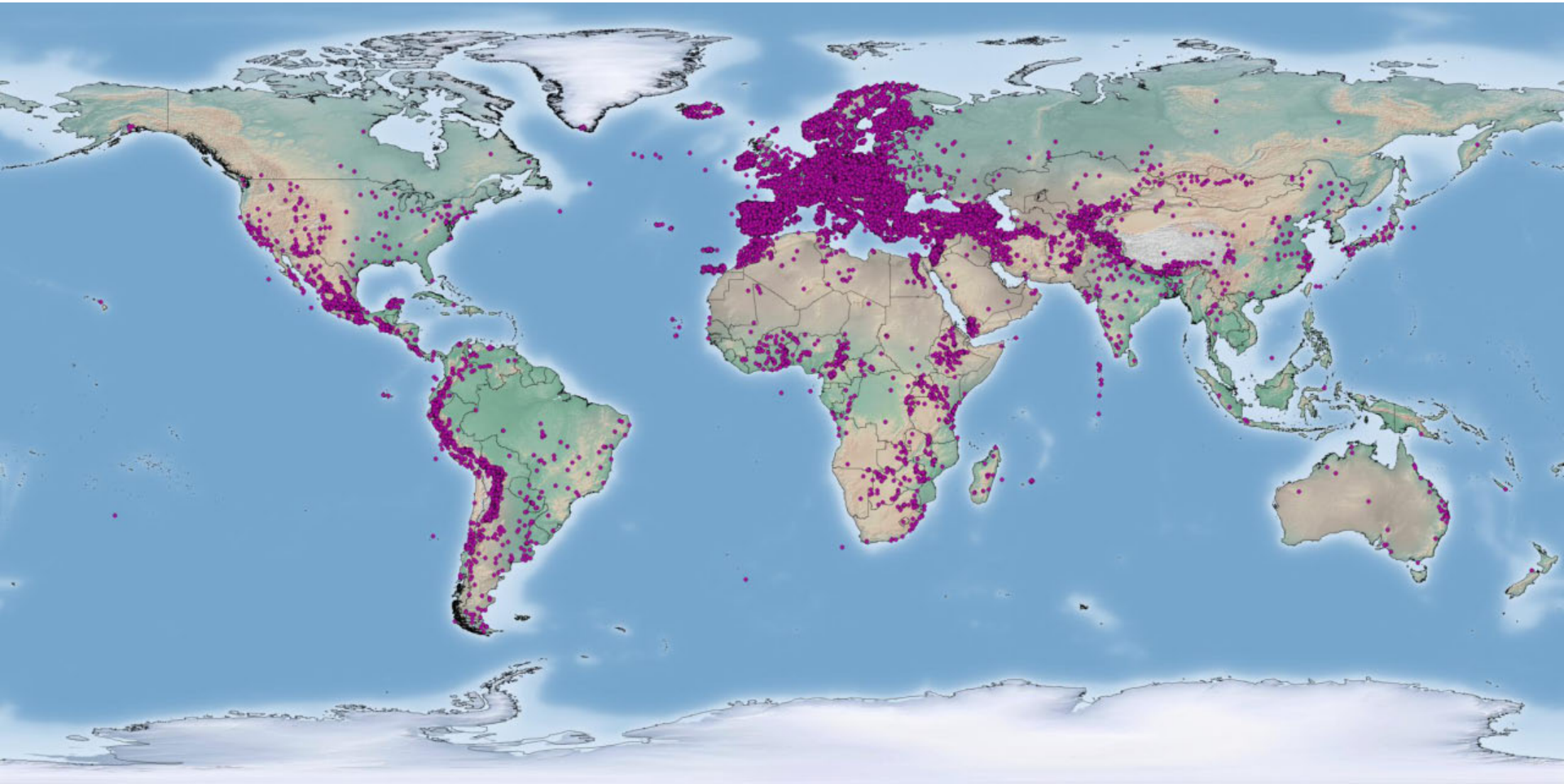
as of 2015-04-17

Countries of origin



as of 2015-04-17

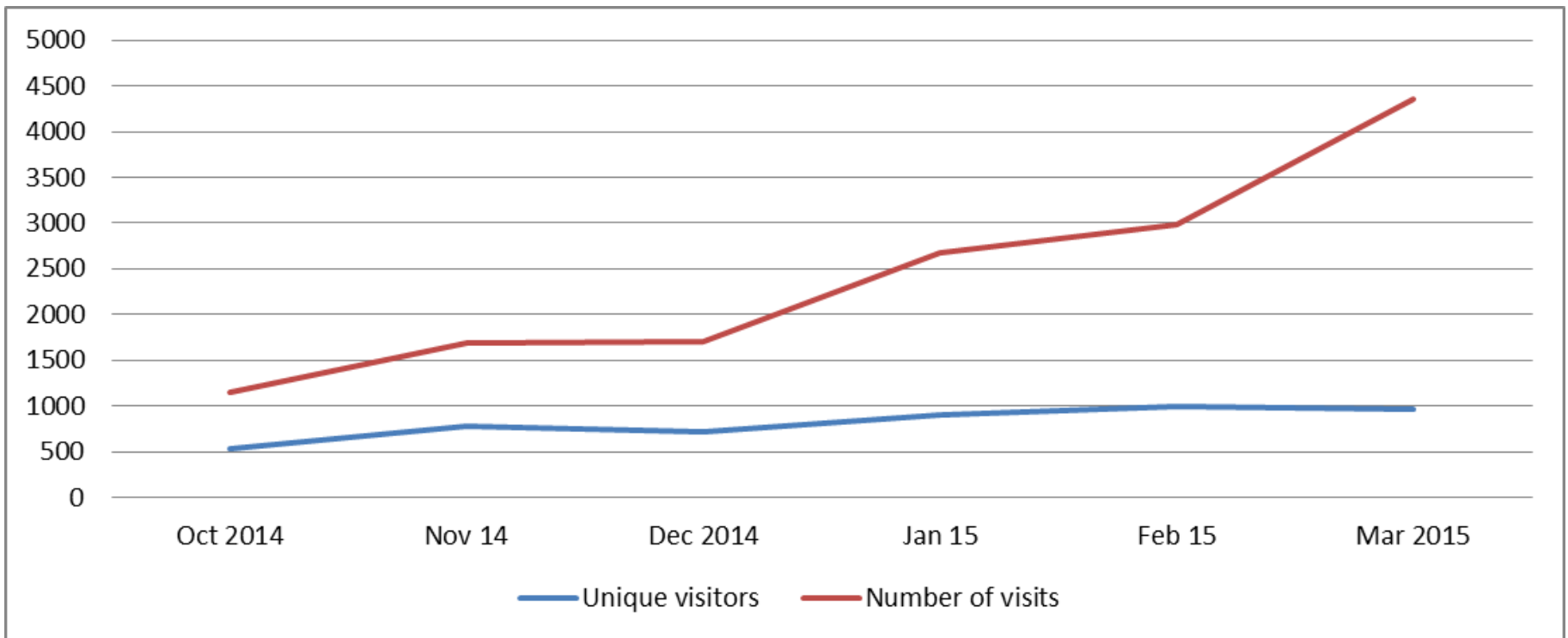
Regions of origin



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

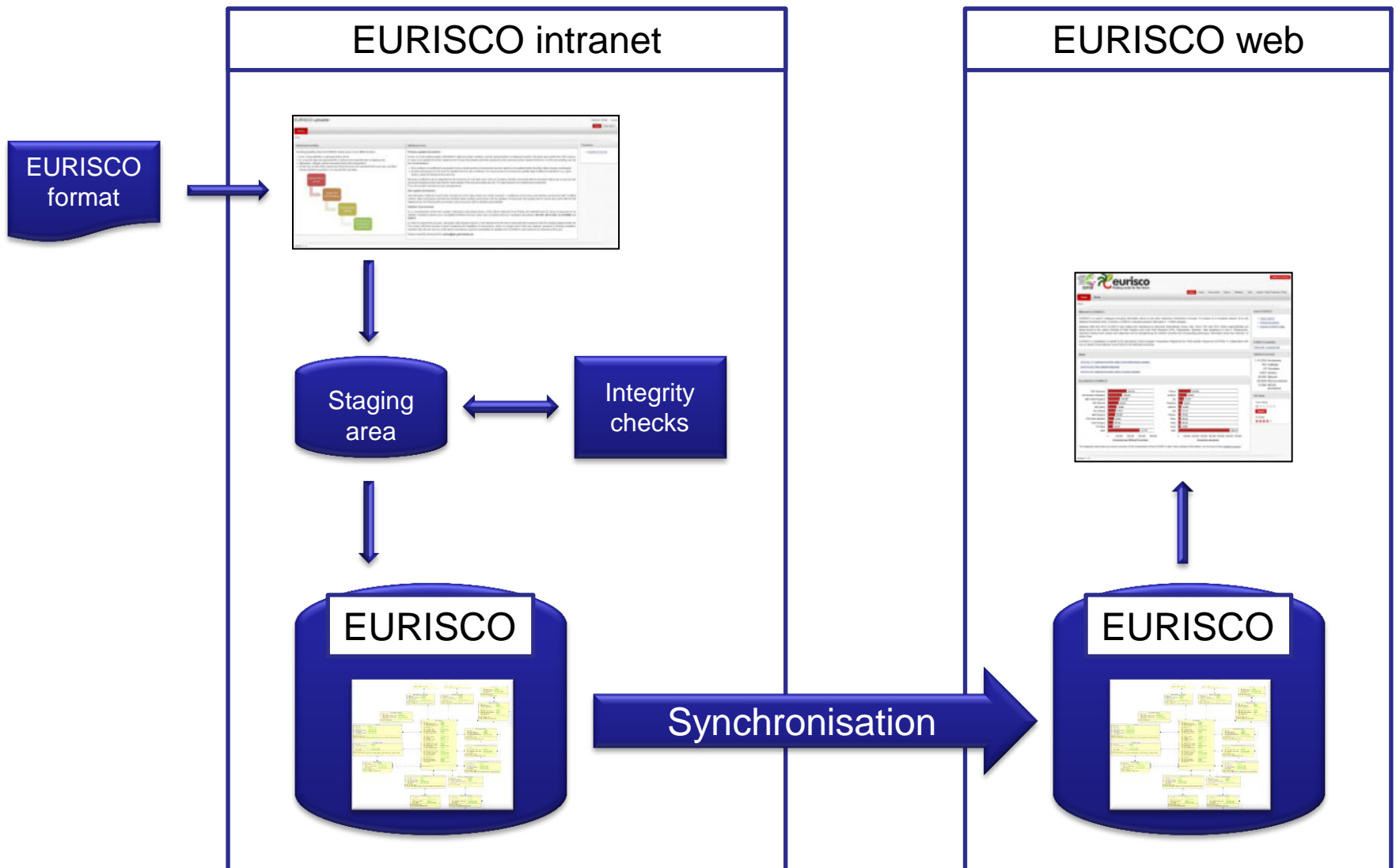
- Preparatory meetings
 - Challenges:
 - Outdated systems
 - Insufficient technical documentation
 - High costs for transfer of as-is status
 - Decisions:
 - No transfer of existing solutions
 - New development from scratch

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



European Cooperative Programme for Plant Genetic Resources
ECP/GR

EURISCO Intranet

Home
About
Documents
Search
Statistics
Links
Imprint / Data Protection Policy

Home News

Home

Welcome to EURISCO

EURISCO is a search catalogue providing information about ex situ plant collections maintained in Europe. It is based on a European network of ex situ National Inventories (NIs). Currently, EURISCO comprises passport data about 1.1 million samples.

Between 2003 and 2014, EURISCO was hosted and maintained by Bioversity International, Rome, Italy. Since 15th April 2014, these responsibilities are being moved to the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. After adapting to a new IT infrastructure, important medium-term actions and objectives will be strengthening the network activities and incorporating phenotypic information about the collection, to name a few.

EURISCO is maintained on behalf of the Secretariat of the European Cooperative Programme for Plant Genetic Resources (ECPGR), in collaboration with and on behalf of the National Focal Points for the National Inventories.

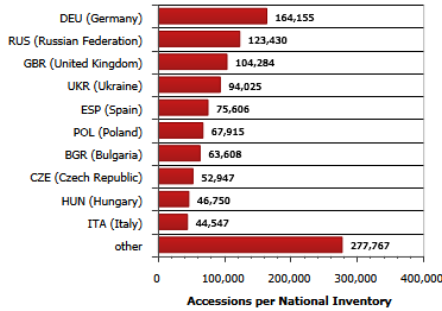
News

[2015-02-17: National Inventory data of the Netherlands updated](#)

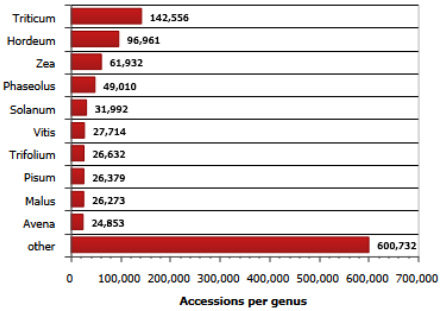
[2015-01-26: Filter options improved](#)

[2015-01-20: National Inventory data of Austria updated](#)

Accessions in EURISCO



Country	Accessions
DEU (Germany)	164,155
RUS (Russian Federation)	123,430
GBR (United Kingdom)	104,284
UKR (Ukraine)	94,025
ESP (Spain)	75,606
POL (Poland)	67,915
BGR (Bulgaria)	63,608
CZE (Czech Republic)	52,847
HUN (Hungary)	46,750
ITA (Italy)	44,547
other	277,767



Genus	Accessions
Triticum	142,556
Hordeum	96,961
Zea	61,932
Phaseolus	49,010
Solanum	31,992
Vitis	27,714
Trifolium	26,632
Pisum	26,379
Malus	26,273
Avena	24,853
other	600,732

The diagrams above give an actual overview of the composition of the EURISCO data. More detailed information can be found at the [statistics section](#).

Search EURISCO

Quick search

Advanced search

Export EURISCO data

EURISCO newsletter

[Subscribe / unsubscribe](#)

Statistical overview

1,115,034 Accessions

362 Institutes

43 Countries

5,657 Genera

36,556 Species

324,049 MLS accessions

12,498 AEGIS accessions

Site rating

Your rating:


☹️ ⭐️ ⭐️ ⭐️ ⭐️ ⭐️

Submit

Average:


⭐️ ⭐️ ⭐️ ⭐️ ⭐️

release 1.1.8



eurisco
Finding seeds for the future

LEIBNIZ INSTITUTE OF PLANT GENETICS AND CROP PLANT RESEARCH



IPK
GATERSLEBEN

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

Search Close

%

Afghanistan

Albania

Algeria

Andorra

Angola

Anguilla

Antarctica

Antigua and Barbuda

Argentina

Armenia

Australia

Austria

Azerbaijan

Bahamas

Search results

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 > National inventory report accession

National inventory report

1-15 16-19

National Inventory	No Of Accessions
Azerbaijan	3
Belarus	2
Belgium	1
Bulgaria	95
Czech Republic	95
Germany	1073
Hungary	23
Israel	1
Moldova	3
Netherlands	167
Nordic Countries	167
Poland	103
Romania	19
Russian Federation	1070
Slovakia	10

1-15 16-19
Download

0.30 s

release 1.1.8

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 > National inventory report accession > Accession list accession

Accession list

1-15 16-30 31-45 46-60 61-75 Next Set >

HOLDING	INSTITUTE	ACCESSION NUMBER	ACCESSION NAME	CROP NAME	ACQUISITION DATE	DETAILS
DEU146	TRI	22480	A.D. HINDUKUSCH		2003-09-12	
DEU146	HOR	19067	A.D.HINDUKUSH		2003-05-21	
DEU146	TRI	24936	ADSCHI AMADJA CHAN		2003-10-08	
DEU146	HOR	16455	ADSCHI AMADJA CHAN		2003-04-29	
DEU146	HOR	15695	ADSCHI AMADJA CHAN		2003-04-24	
DEU146	TRI	21884	AFGHAN NO VAWW 409/55		2003-09-09	
DEU146	HOR	17219	AFGHAN.NR.5921		2003-05-07	
DEU146	HOR	17009	AFGHANISTAN 43		2003-05-06	
DEU146	TRI	28643	AFGHANISTANISCHER SOMMERWEIZEN		2003-11-11	
DEU146	TRI	28641	AFGHANISTANISCHER SOMMERWEIZEN		2003-11-11	
DEU146	HOR	15802	AHMEDI DEWANA		2003-04-24	
DEU146	HOR	15800	AHMEDI DEWANA		2003-04-24	
DEU146	HOR	15793	AHMEDI DEWANA		2003-04-24	
DEU146	HOR	15778	APCEI		2003-04-24	
DEU146	TRI	24844	ARANDAN		2003-10-07	

Accession details

The screenshot shows the eurisco website interface. At the top left is the eurisco logo with the tagline "Finding seeds for the future". Below the logo is a navigation bar with "Search", "Advanced search", and "Export EURISCO data". A secondary navigation bar includes "Home", "About", "Documents", "Search", "Statistics", "Links", "Inquiry", and "Data Protection Policy". The main content area is divided into several sections, each with a dropdown arrow and a title: "National inventory", "Holding institute", "Accession", "Taxonomy", "Acquisition/Storage", "Collective", and "Donor". Each section contains specific data points related to the accession.

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 conv. vulgare var. hybernum
Subtaxa Authority Viborg

Acquisition/Storage
Acquisition Date 1922
Acquisition Source Other
Germplasm Storage Short term

Collective
Collecting Number 29
Collecting Institute Code DEU146
Collecting Latitude 47.0457
Collecting Longitude 10.83
Collecting Site St. Leonhard i. Pflztal, Tirol
Description Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research

Donor
Donor Institute Code AUST017
Donor Institute Name Erich Mayr
Description Erich Mayr

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 conv. vulgare var. hybernum
Subtaxa Authority Viborg

...

User specific export

EURISCO Intranet

[Home](#) [About](#) [Documents](#) [Search](#) [Statistics](#) [Links](#) [Imprint / Data Protection Policy](#)

[Search](#) [Advanced search](#) [Export EURISCO data](#)

Home > Search > EURISCO download

Export EURISCO data

Descriptor(s) *

COLLNUMB	NICODE
COLLCODE	INSTCODE
SPAUTHOR	ACCENUMB
SUBTAXA	GENUS
SUBTAUTHOR	SPECIES
CROPNAME	ORIGCTY
ACCENAME	
ACQDATE	
COLLSITE	
LATITUDE	

National Inventory *

Genus

Download full dataset

Downloading the whole dataset of EURISCO causes a very long page loading time. Thus, for performance reasons data from only one National Inventory can be downloaded at once. A precalculated dump of the whole dataset (in EURISCO format) can be downloaded here:

[EURISCO dump](#)
28.66 MB
Created: 2015-02-18

Filtered values

1-5 [6-10](#) [11-15](#) [16-20](#) [21-25](#) [Next Set >](#)

NICODE	INSTCODE	ACCENUMB	GENUS	SPECIES	ORIGCTY
ALB	ALB017	AGB1404	Aegilops	ovata	ALB
ALB	ALB017	AGB1403	Aegilops	ovata	ALB
ALB	ALB017	AGB1402	Aegilops	ovata	ALB
ALB	ALB017	AGB1401	Aegilops	ovata	ALB
ALB	ALB017	AGB1717	Aegilops	ovata	ALB

1-5 [6-10](#) [11-15](#) [16-20](#) [21-25](#) [Next Set >](#)

0.01 s

release 1.1.8

Statistics

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

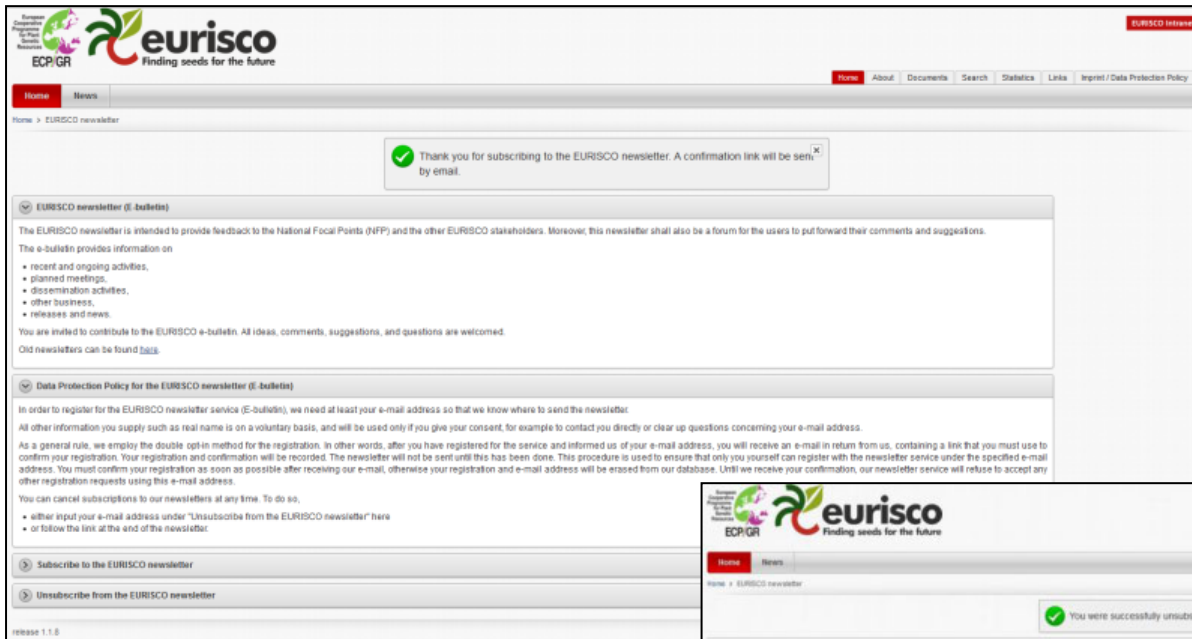
National Inventory	Accessions	Percentage
DEU (Germany)	38,670	3.47%
NLD (Netherlands)	22,705	2.04%
UKR (Ukraine)	19,687	1.77%
NGB (Nordic Countries)	12,491	1.12%
ROU (Romania)	6,777	0.61%
ITA (Italy)	5,977	0.54%
SVK (Slovakia)	2,883	0.26%
ISR (Israel)	2,849	0.26%
CZE (Czech Republic)	2,266	0.20%
AUT (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 >](#)

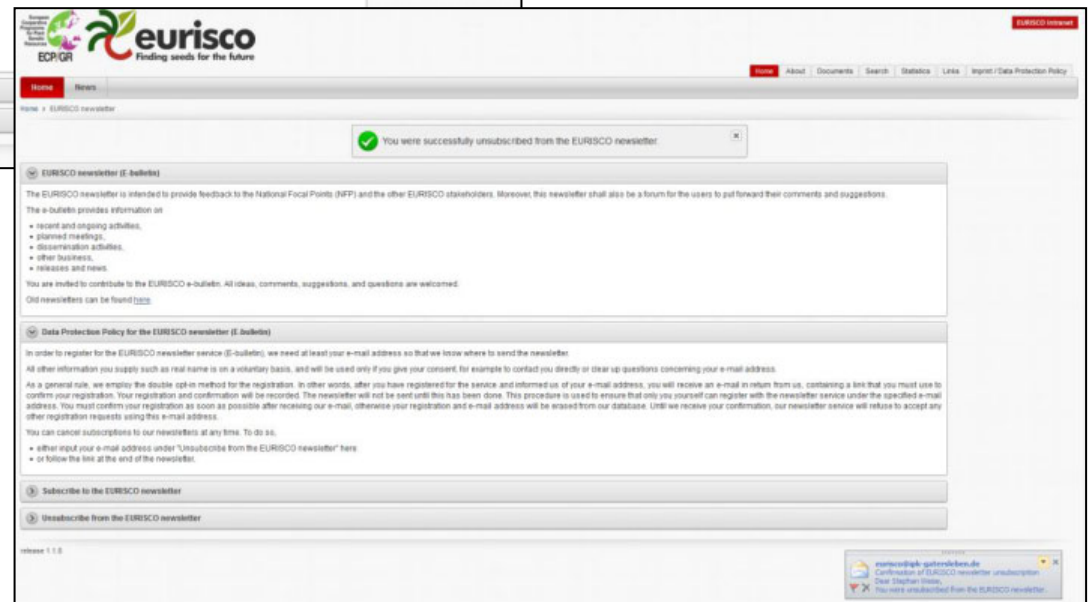
Search EURISCO

- [Quick search](#)
- [Advanced search](#)
- [Export EURISCO data](#)

Newsletter subscription



The screenshot shows the EURISCO website interface. At the top left is the EURISCO logo with the tagline "Finding seeds for the future" and "ECP GR". A navigation menu includes "Home", "About", "Documents", "Search", "Statistics", "Links", and "Imprint / Data Protection Policy". A secondary menu has "Home" and "News". A central notification box with a green checkmark states: "Thank you for subscribing to the EURISCO newsletter. A confirmation link will be sent by email." Below this, there are sections for "EURISCO newsletter (E-bulletin)", "Data Protection Policy for the EURISCO newsletter (E-bulletin)", and "Subscribe to the EURISCO newsletter". The "Subscribe" section includes a link to "Unsubscribe from the EURISCO newsletter". The footer indicates "Issue 1.1.8".



The screenshot shows the EURISCO website interface after unsubscription. The navigation and menu elements are identical to the previous screenshot. A central notification box with a green checkmark states: "You were successfully unsubscribed from the EURISCO newsletter." Below this, the "EURISCO newsletter (E-bulletin)" and "Data Protection Policy for the EURISCO newsletter (E-bulletin)" sections are visible. The "Subscribe to the EURISCO newsletter" and "Unsubscribe from the EURISCO newsletter" links are still present. The footer indicates "Issue 1.1.8".

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 uploader

Welcome: WEISE Logout

Home Data import

Home

General proceeding

Inserting/updating data into EURISCO takes place in four different steps:

- First, a new data file is uploaded to the server.
- In a second step, the uploaded file is parsed and imported into a staging area.
- Afterwards, integrity checks are performed in the background.
- At last, the results of the import and check process are reported to the user who can then decide whether to publish or to discard the new data.

Upload file to server

Import file into database

Perform data checks

Decide about publishing or withdrawal

Update process

Previous update mechanism

So far, no incremental updates of EURISCO data took place. Instead, only full replacements on National Inventory (NI) level were performed. This means, in case of an update the whole dataset of an NI was first deleted and then replaced by the reworked whole dataset of this NI. For this proceeding, we see two disadvantages:

- This solution is not efficient, especially if only a small number of accession records needs to be updated while the other data remains unchanged.
- All data belonging to an NI must be updated at once (all or nothing). It is not possible to successively update data of different collections (e.g. gene banks), which all belong to the same NI.

Moreover, EURISCO will be extended for the inclusion of C-E data soon. Since C-E data is directly connected with the passport data of an accession, the above proceeding would imply that for each update of the passport data also all C-E data needs to be deleted and reimported. Thus, the update mechanism was reengineered.

New update mechanism

Henceforward, National Focal Points should only send data, which was really changed — additional accessions and existing accessions with modified content. New accessions will then be inserted while existing accessions will be updated. If necessary, the update will of course also work with the full dataset of an NI. Following this procedure, no accessions will be deleted automatically.

Deletion of accessions

As a consequence of the new update mechanism described above, in the future National Focal Points will explicitly have to name accessions to be deleted. Therefore, please use a simplified EURISCO format, which only comprises the four mandatory descriptors **NCODE**, **INSTCODE**, **ACCENUMB** and **GENUS**.

In order to support this process, during the data integrity checks a new dataset of an NI will be automatically compared with the existing dataset of this NI. The system will then provide a report containing the identifiers 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 by the user.

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

Templates

EURISCO format

release 1.1.1

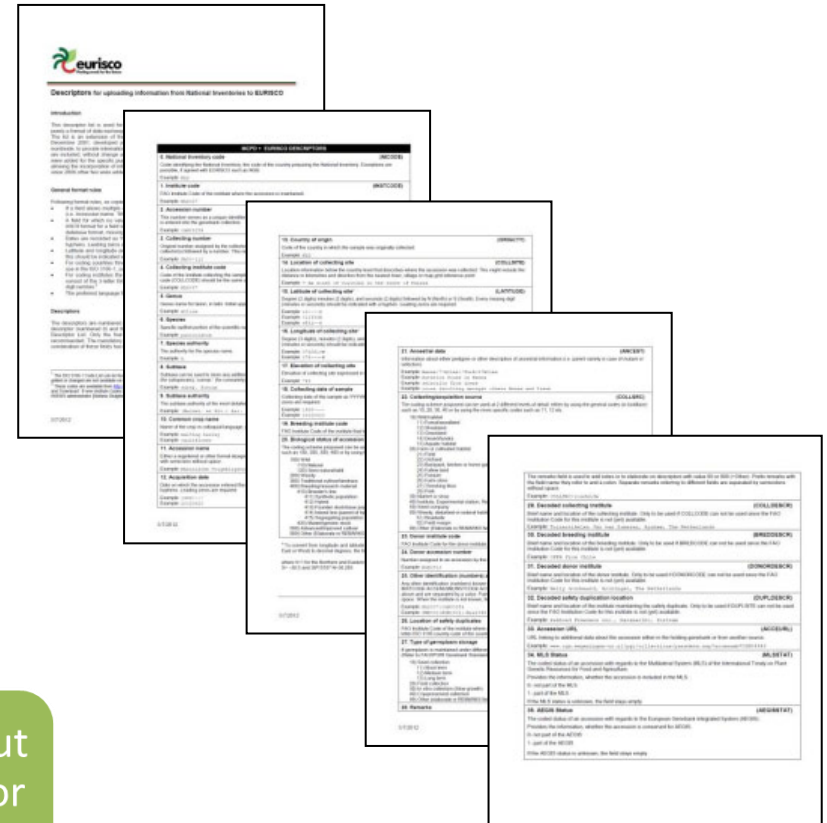
EURISCO intranet II

Upload file to server

Import file into database

Perform data checks

Decide about publishing or withdrawal



File upload

- Text file
- Tab separated
- UTF-8

EURISCO uploader

Welcome: WEISE Logout

Home Data import

Upload file Import file Integrity check results Decision about update

Home > Upload file

First step: File upload

The first step of importing new or modified data into EURISCO is to upload a file containing National Inventory data to the EURISCO server. This file must be formatted in accordance with the MCPD EURISCO format. It must be a UTF-8 encoded text file with tabulator-separated columns.

Select file to upload

Please select a text file (.txt) in tab-delimited format:

File *

release 1.1.2

...Uploading file...

File upload

File import

Integrity checks

Final decision

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 parse the data. Please select the previously uploaded file from the popup list, select a validation profile. Afterwards, necessary integrity checks will be performed on the imported data. You will be informed of the check results by email.

Import settings

Import File *

Validation Profile *

Notification Email *

release 1.1.2

Validation Profile ✕

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.

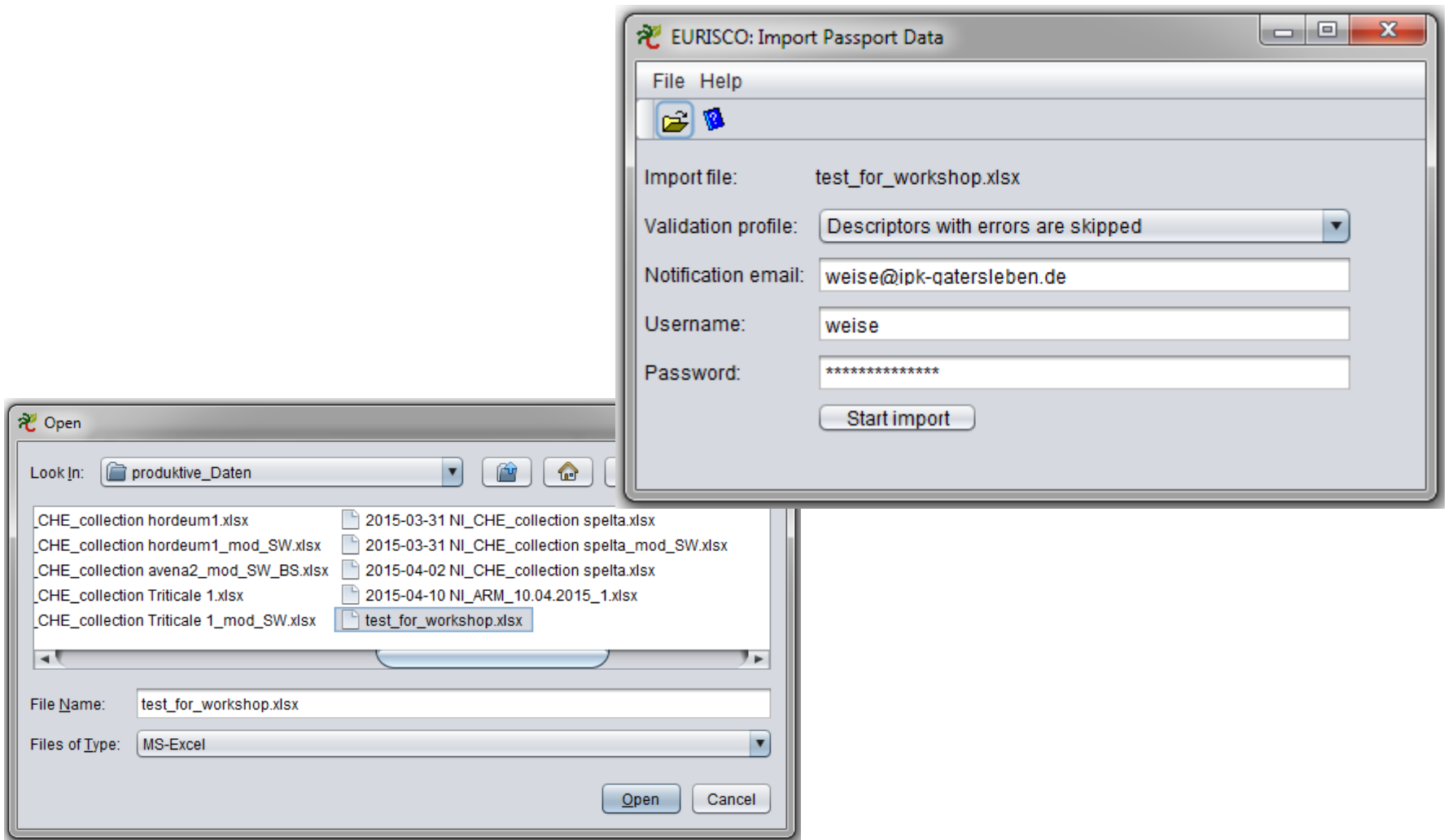
Descriptors with errors are skipped:

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.

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

EURISCO uploader Welcome: WEISE Logout

[Home](#) [Data import](#)

[Upload file](#) [Import file](#) **Integrity check results** [Decision about update](#)

Home > EURISCO uploader Welcome: WEISE Logout

[Home](#) [Data import](#)

[Upload file](#) [Import file](#) **Integrity check results** [Decision about update](#)

Home > EURISCO uploader Welcome: WEISE Logout

[Home](#) [Data import](#)

[Upload file](#) [Import file](#) **Integrity check results** [Decision about update](#)

Home > Upload file > Import file > Check results overview

Third step: Data 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 (setup finished, import running, import finished). On the sub-pages, all errors will be listed (grouped by descriptor).

National Inventory	Filename	File uploaded	Notification email	End of import	Import status
AUT	test_for_workshop.txt	2015-04-22 13:46:30	weise@ipk-gatersleben.de	2015-04-22 14:10:16	Import finished

release 1.1.2

```
graph TD; A[File upload] --> B[File import]; B --> C[Integrity checks]; C --> D[Final decision];
```

Integrity checks II – error report

EURISCO uploader Welcome: WEISE Logout

[Home](#) [Data import](#)

[Upload file](#) [Import file](#) **[Integrity check results](#)** [Decision about update](#)

[Home](#) > [Upload file](#) > [Import file](#) > [Check results overview](#) > [Errors per descriptor](#)

Errors per descriptor

Search:

Actions

1 - 14

Descriptor	Number Of Errors
ACQDATE	44
BREDCODE	2
COLLDATE	4
COLLSRC	21
DONORCODE	21
DUPLSITE	45
ELEVATION	2
GENUS	1
LATITUDE	8
LONGITUDE	22
MLSSTAT	1
ORIGCTY	1
REMARKS	6
STORAGE	3

1 - 14

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

```
graph TD; A[File upload] --> B[File import]; B --> C[Integrity checks]; C --> D[Final decision];
```

release 1.1.2

Integrity checks III - example

- Wrong date format

EURISCO uploader

Welcome: WEISE Logout

Home Data import

Upload file Import file **Integrity check results** Decision about update

Home > Upload file > Import file > Check results overview > Errors per descriptor > Error details

1 - 15

Descriptor	Error Description
ACQDATE	Line 2201: ACQDATE 2002-- invalid.
ACQDATE	Line 2202: ACQDATE 2002-- invalid.
ACQDATE	Line 2203: ACQDATE 2002-- invalid.
ACQDATE	Line 2162: ACQDATE 2002-- invalid.
ACQDATE	Line 2163: ACQDATE 2002-- invalid.
ACQDATE	Line 2164: ACQDATE 2002-- invalid.
ACQDATE	Line 2220: ACQDATE 2002-- invalid.
ACQDATE	Line 2221: ACQDATE 2002-- invalid.
ACQDATE	Line 2218: ACQDATE 2002-- invalid.
ACQDATE	Line 2222: ACQDATE 2002-- invalid.
ACQDATE	Line 2171: ACQDATE 2002-- invalid.
ACQDATE	Line 2172: ACQDATE 2002-- invalid.
ACQDATE	Line 2192: ACQDATE 2002-- invalid.
ACQDATE	Line 2219: ACQDATE 2002-- invalid.
ACQDATE	Line 2173: ACQDATE 2002-- invalid.

1 - 15

release 1.1.2

File upload
↓
File import
↓
Integrity checks
↓
Final decision

Integrity checks IV - example

- Invalid or multiple donor codes

EURISCO uploader

Welcome: WEISE Logout

Home Data import

Upload file Import file **Integrity check results** Decision about update

Home > Upload file > Import file > Check results overview > Errors per descriptor > Error details

Q- Go Actions

1 - 15

Descriptor	Error Description
DONORCODE	Line 647: DONORCODE FAO031 invalid.
DONORCODE	Line 648: DONORCODE FAO031 invalid.
DONORCODE	Line 649: DONORCODE FAO031 invalid.
DONORCODE	Line 806: DONORCODE ROU006 invalid.
DONORCODE	Line 8429: DONORCODE DEU146 / GBR040 invalid.
DONORCODE	Line 8512: DONORCODE AUT016/Versailles invalid.
DONORCODE	Line 8527: DONORCODE AUT016/INIA invalid.
DONORCODE	Line 8851: DONORCODE DEU146 / DEU235 invalid.
DONORCODE	Line 8853: DONORCODE DEU146 / USA126 invalid.
DONORCODE	Line 11057: DONORCODE DEU146 / DEU040 invalid.
DONORCODE	Line 11068: DONORCODE DEU146 / CHE008 invalid.
DONORCODE	Line 11069: DONORCODE DEU146 / PRT008 invalid.
DONORCODE	Line 11070: DONORCODE DEU146 / DEU032 invalid.
DONORCODE	Line 11071: DONORCODE DEU146 / SVN006 invalid.
DONORCODE	Line 11072: DONORCODE DEU146 / DEU152 invalid.

1 - 15

release 1.1.2

```
graph TD; A[File upload] --> B[File import]; B --> C[Integrity checks]; C --> D[Final decision];
```

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

Q Go Rows 15 Actions

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

EURISCO uploader

Welcome: WEISE Logout

Home Data import

Upload file Import file Integrity check results **Decision about update**

Home > Upload file > Import file > Decision about update > Final decision

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

File upload

File import

Integrity checks

Final decision

release 1.1.2

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 resolutions.

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 accessions¹. 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

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



<http://www.ars-grin.gov>

Catalogue of Life



<http://www.catalogueoflife.org>

Increase data quantity/quality III

- Data quality
 - Improve location data quality

▼ Taxonomy

Genus **Medicago**
Species **murex**

▼ Acquisition/storage

Acquisition Source **Roadside**

▼ Collection

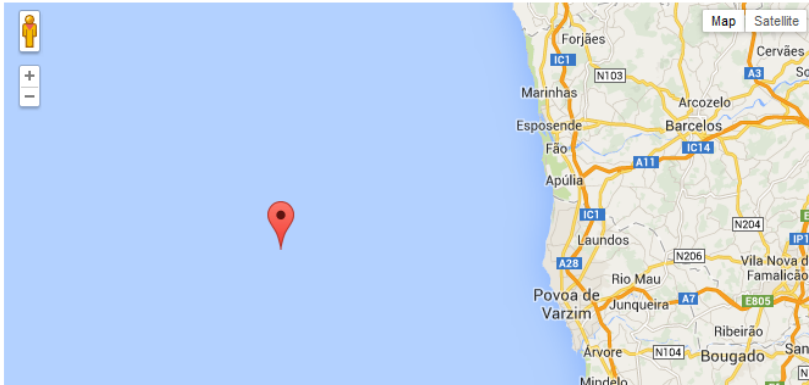
Collecting Institute Code

Collecting Date **1987**

Collecting Latitude **41.4264**

Collecting Longitude **-9.0831**

Collecting Elevation **635**



Increase data quantity/quality IV

- Data quality
 - Reduce inconsistencies

Quick Search

Taxonomy Accession Status

Genus: Contains R.da

Genus containing "R.damascena" ✖

- ?? ?????.?????.R.damascena
- R.damascena x R.gallica

Name	Acquisition Date	Details
	01/01/0931	

Collecting Elevation	2053
Collecting Locality	""Shirak. Torosgyugh;near ""

Increase data quantity/quality V

- Data quality
 - Fix encoding issues

The screenshot displays a data entry interface with several fields containing encoding artifacts. Red circles highlight these issues:

- Accession Name:** The field contains "nagykalló, 47°52'n, 21°51'e". The coordinates "47°52'n, 21°51'e" are circled in red.
- Accession Name (dropdown):** A list of accession names is shown, including "OMAR", "ELMAR", "öäë-800", "öäë-135", and "öäë-798". The entries "öäë-800", "öäë-135", and "öäë-798" are circled in red.
- Collecting Locality:** The field contains ", Slobozia Mare, Region Vulcanesti, Slobozia Mare". The word "Vulcöneöti" is circled in red.
- Collecting Elevation:** The field is empty.
- Collecting Locality (bottom):** The field contains a long string of "ö" characters, circled in red.

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

Additional 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**