# What can a C&E module in EURISCO give us?

Forage and Barley ECPGR workshop, 14-16 March 2017, Malmö, Sweden

Stephan Weise
14 March 2017

# BACKGROUND

# C&E data I

- Previously not available in EURISCO
- Of high importance to users of PGR data
  - Determines value of germplasm for breeding and research
- Difficult to handle due to lots of “standards”
  - Different descriptor names/synonyms
  - Different rating scales
    - Nominal, ordinal, metric scale
  - Different amounts of meta information
    - When, where, how, by whom?
    - Experiment set-up, treatment etc.

# C&E data II

- Discussed in detail in previous years
- Suitable proposal compiled by the ECPGR Doc&Info Network in 2009
  - As simple as possible
  - As few fields as possible

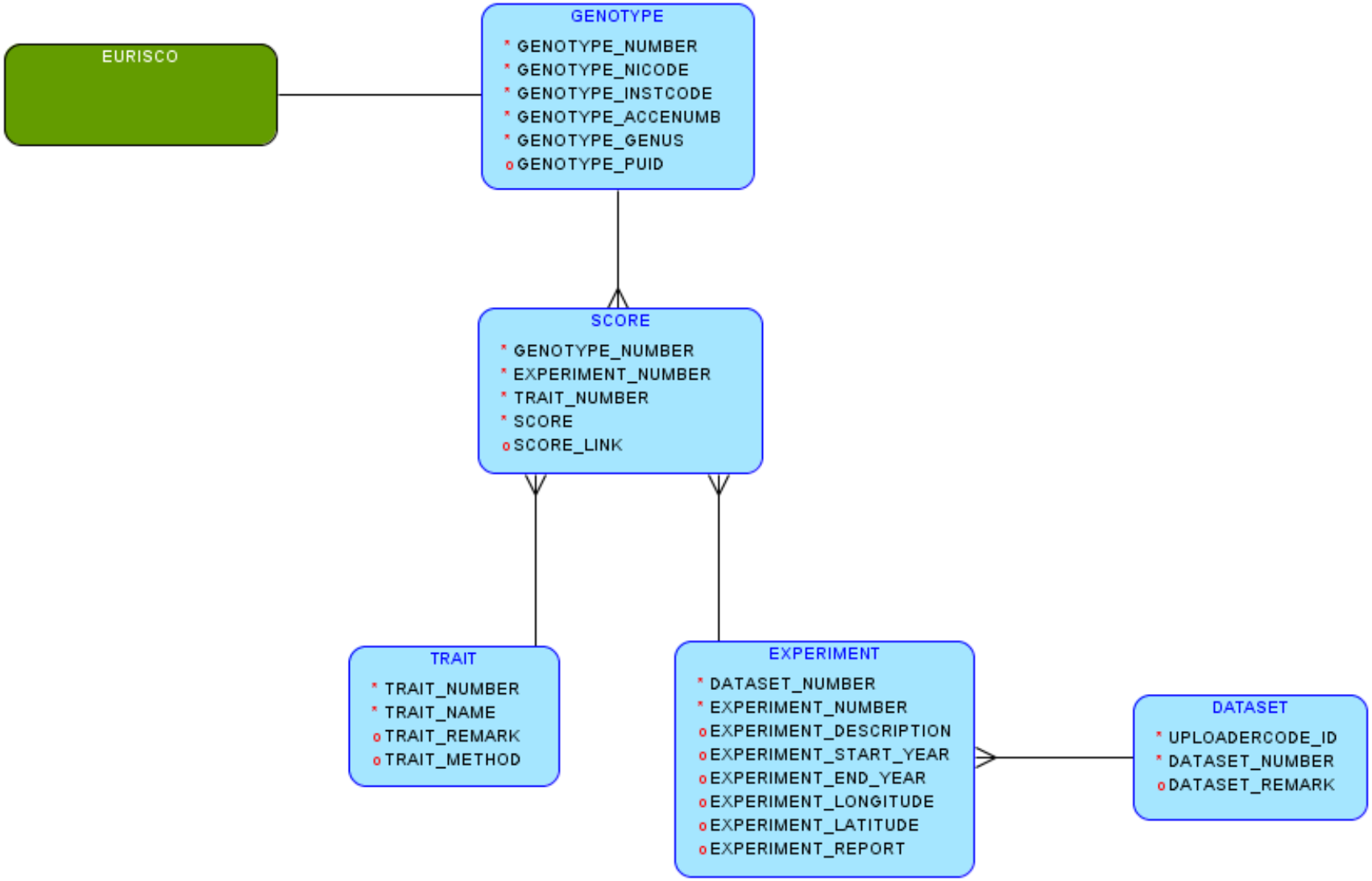
    → “minimum consensus”
- D&I meeting 2014
  - Pragmatic approach:
    - No standardisation of trait, scale or experimental design
    - Only standardisation of exchange format (as simple as possible)
    - Import of existing data as-is → reach critical mass
    - Use proposal of 2009 (+ small adjustments)

# DATA MODEL

# Data model for C&E data I

- Five components
  - **GENOTYPE**
    - Identified by EURISCO descriptors
  - **DATASET**
    - May comprise different experiments
  - **EXPERIMENT**
    - Multiple genotypes are scored for different traits
  - **TRAIT**
    - Characteristic feature to be scored
  - **SCORE**
    - Value of a trait for a genotype

# Data model for C&E data II



# TEMPLATES

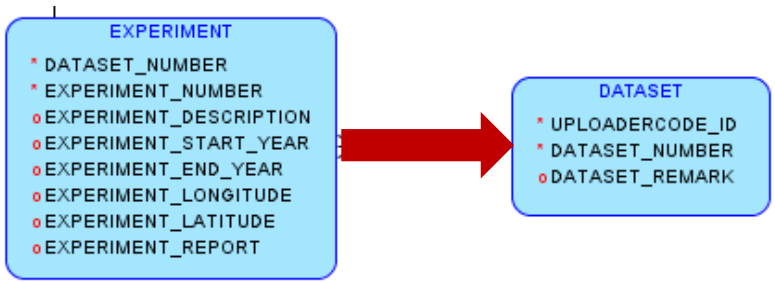
# DATASET

- Enables to upload multiple experiments at once
- Fields:
  - **UPLOADERCODE***:
    - ID of registered authorised data provider
    - Provided by EURISCO
  - **DATASET_NUMBER***:
    - To link experiments with datasets
    - Unique and persistent for the data provider
  - DATASET_REMARK:
    - General remark for all scores in the dataset

| UPLOADERCODE | DATASET_NUMBER | DATASET_REMARK |
|---|---|---|
| DEU | 1 | This dataset contains forage grass accessions of IPK. |
| … | … | … |
| … | … | … |

# EXPERIMENT I

- Meta data helping to interpret C&E data
  - Experiment set-up
  - Weather conditions
  - Soil conditions
  - Experiment location
  - …
- Fields:
  - **DATASET_NUMBER***:
    - Reference of the dataset
  - **EXPERIMENT_NUMBER***:
    - To link scores with experiments
    - Unique and persistent for the data provider



# EXPERIMENT II

- Fields (cont.):
  - EXPERIMENT_DESCRIPTION:
    - Brief English description
    - Information necessary for interpreting the scores, e.g. set-up
  - EXPERIMENT_START_YEAR:
    - Year in which the experiment was performed (or started)
  - EXPERIMENT_END_YEAR:
    - Year in which the experiment ended
  - EXPERIMENT_LONGITUDE:
    - Longitude of experimental site
  - EXPERIMENT_LATITUDE:
    - Latitude of experimental site
  - EXPERIMENT_REPORT:
    - Reference to a report (either report file or report URL)

# EXPERIMENT III

| DATASET_NUMBER | EXPERIMENT_NUMBER | EXPERIMENT_DESCRIPTION | EXPERIMENT_START_YEAR | EXPERIMENT_END_YEAR | EXPERIMENT_LONGITUDE | EXPERIMENT_LATITUDE | EXPERIMENT_REPORT |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Characterisation data of Lolium perenne | 1999 | 2000 | 11.278414 | 51.826059 | http://... |
| 1 | 2 | Characterisation data of Lolium perenne | 2000 | 2001 | 11.278414 | 51.826059 | http://... |
| 1 | 3 | Characterisation data of Lolium perenne | 2001 | 2002 | 11.278414 | 51.826059 | http://... |
| 1 | 4 | Evaluation data of Lolium perenne (4 replications per accession) | 2002 | 2005 | 11.278414 | 51.826059 | http://... |
| ... | ... | ... | ... | | ... | ... | ... |

# TRAIT I

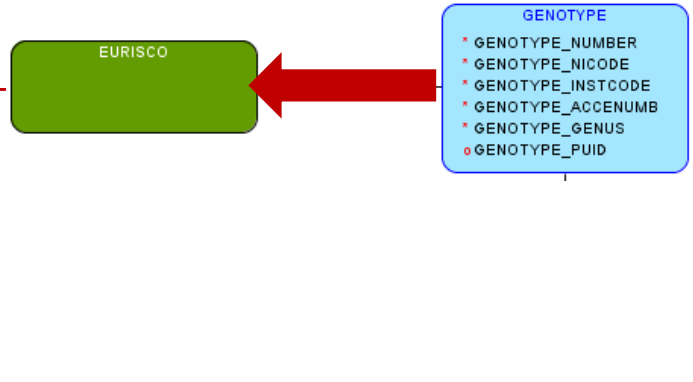
- Describes phenotypic traits and the methods used for scoring
- Fields:
  - **TRAIT_NUMBER***:
    - Unique, temporary number of the trait in the dataset
  - **TRAIT_NAME***:
    - English name of the trait
  - TRAIT_REMARK:
    - General remark helping to interpret the trait
  - TRAIT_METHOD:
    - English description of the used method + scale

# TRAIT II

| TRAIT_NUMBER | TRAIT_NAME | TRAIT_REMARK | TRAIT_METHOD |
|---|---|---|---|
| 1 | Sowing date | … | Date |
| 2 | Emerging date | Date, when first shoot is visible | Date |
| 3 | Growing before winter | Quality of plant growing before winter | Rating value from 1 (min) – 9 (max) |
| 4 | Stem height min | In flowering time, the shortest plant | Measurement [cm] |
| … | … | … | … |

# GENOTYPE I

- All accessions for which C&E data will be uploaded
- Fields:
  - **GENOTYPE_NUMBER***:
    - Unique, temporary number of the genotype in the dataset
  - **GENOTYPE_NICODE***:
    - National Inventory code from EURISCO
  - **GENOTYPE_INSTCODE***:
    - Holding institute code from EURISCO
  - **GENOTYPE_ACCENUMB***:
    - Accession number from EURISCO
  - **GENOTYPE_GENUS***:
    - Genus from EURISCO
  - GENOTYPE_PUID:
    - Placeholder for a permanent unique identifier (DOI → ITPGRFA)

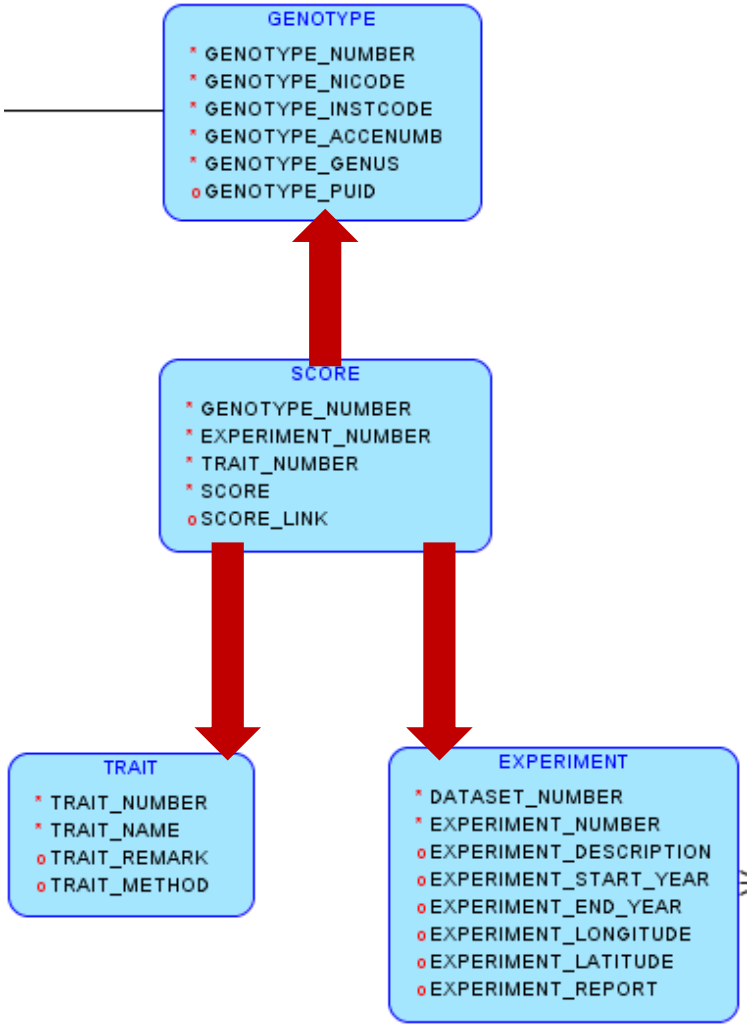


# GENOTYPE II

| GENOTYPE_NUMBER | GENOTYPE_NICODE | GENOTYPE_INSTCODE | GENOTYPE_ACCENUMB | GENOTYPE_GENUS | GENOTYPE_PUID |
|---|---|---|---|---|---|
| 1 | DEU | DEU271 | GR 142 | Lolium | |
| 2 | DEU | DEU271 | GR 476 | Lolium | |
| 3 | DEU | DEU271 | GR 550 | Lolium | |
| 4 | DEU | DEU271 | GR 2670 | Lolium | |

# SCORE I

- Observed phenotypic values of the accessions
- Fields:
  - **GENOTYPE_NUMBER***:
    - Reference to a genotype
  - **EXPERIMENT_NUMBER***:
    - Reference to an experiment
  - **TRAIT_NUMBER***:
    - Reference to a trait
  - **SCORE***:
    - Observed score
  - SCORE_LINK:
    - Link to a publication on accession level



# SCORE II

| GENOTYPE_NUMBER | EXPERIMENT_NUMBER | TRAIT_NUMBER | SCORE | SCORE_LINK |
|---|---|---|---|---|
| 1 | 1 | 1 | 19990313 | http://... |
| 1 | 1 | 3 | 7 | http://... |
| 4 | 4 | 1 | 20020401 | … |
| 4 | 4 | 4 | 21 | http://... |
| … | … | … | … | … |
| … | … | … | … | … |

# Putting all this together

| GENOTYPE_NUMBER | GENOTYPE_NICODE | GENOTYPE_INSTCODE | GENOTYPE_ACCENUMB | GENOTYPE_GENUS | GENOTYPE_PUID |
|---|---|---|---|---|---|
| 1 | DEU | DEU271 | GR 142 | Lolium | |
| 2 | DEU | DEU271 | GR 476 | Lolium | |
| 3 | DEU | DEU271 | GR 550 | Lolium | |
| 4 | DEU | DEU271 | GR 2670 | Lolium | |

GENOTYPE

TRAIT

| TRAIT_NUMBER | TRAIT_NAME | TRAIT_REMARK | TRAIT_METHOD |
|---|---|---|---|
| 1 | Sowing date | … | Date |
| 2 | Emerging date | … | Date |
| 3 | Growing before winter | … | Rating value from 1 (min) – 9 (max) |
| 4 | Stem height min | In flowering time, the shortest plant | Measurement [cm] |
| … | … | … | … |

| GENOTYPE_NUMBER | EXPERIMENT_NUMBER | TRAIT_NUMBER | SCORE | SCORE_LINK |
|---|---|---|---|---|
| 1 | 1 | 1 | 19990313 | http://… |
| 1 | 1 | 3 | 7 | http://… |
| 4 | 4 | 1 | 20020401 | … |
| 4 | 4 | 4 | 21 | http://… |
| … | … | … | … | … |
| … | … | … | … | … |

SCORE

| DATASET_NUMBER | EXPERIMENT_NUMBER | EXPERIMENT_DESCRIPTION | EXPERIMENT_START_YEAR | EXPERIMENT_END_YEAR | EXPERIMENT_LONGITUDE | EXPERIMENT_LATITUDE | EXPERIMENT_REPORT |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Characterisation data of Lolium perenne | 1999 | | 11.278414 | 51.826059 | http://… |
| 1 | 2 | Characterisation data of Lolium perenne | 2000 | | 11.278414 | 51.826059 | http://… |
| 1 | 3 | Characterisation data of Lolium perenne | 2001 | | 11.278414 | 51.826059 | http://… |
| 1 | 4 | Evaluation data of Lolium perenne (4 replications per accession) | 2002 | 2005 | 11.278414 | 51.826059 | http://… |
| … | … | … | … | | … | … | … |

EXPERIMENT

| UPLOADERCODE | DATASET_NUMBER | DATASET_REMARK |
|---|---|---|
| deu_uploader01 | 1 | This dataset contains forage grass accessions. |
| … | … | … |
| … | … | … |

DATASET

# PROCEEDING

# Proceeding for data upload I

- Prerequisite:
  - Only non-confidential C&E data
  - Only data of accessions listed in EURISCO

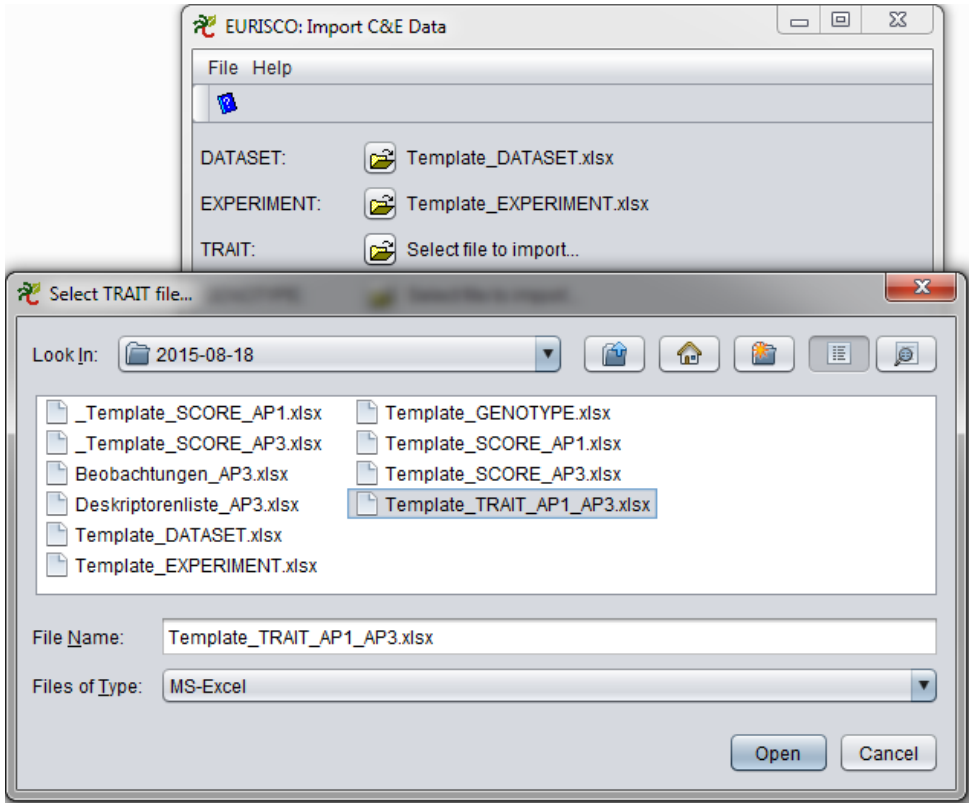
- Impact
  - NFPs responsible for data upload (Data Sharing Agreements)
    →May nominate users for (sub) accounts for data uploads
    →NFPs must approve data before publication

- Data formatting
  - According to the templates in MS Excel (.xlsx) files
  - One file for each template

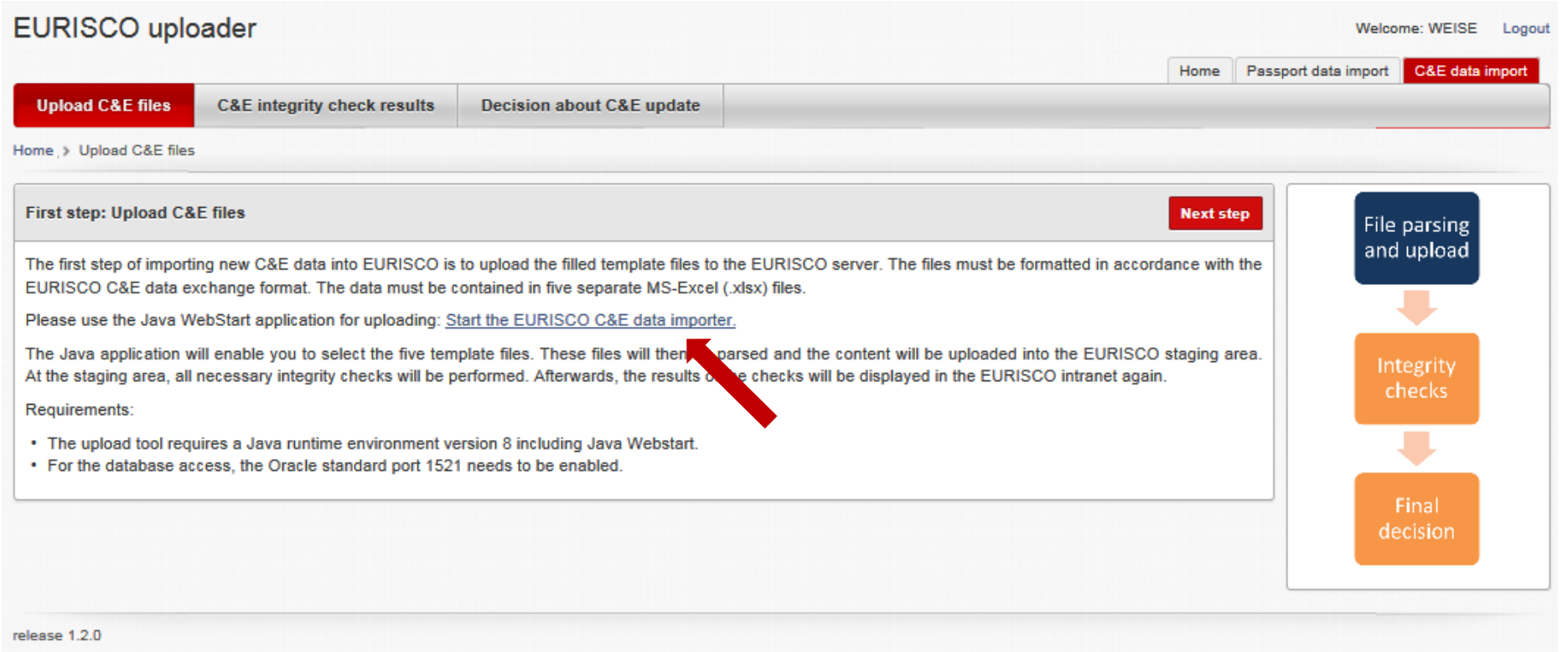
- Upload via EURISCO intranet

# Proceeding for data upload II

- Three steps
  1. File parsing and upload via Java tool *(data owner / NFP)*
  2. Data integrity checks *(EURISCO management)*
  3. Approval / withdrawal of data for publishing on the EURISCO website *(NFP)*

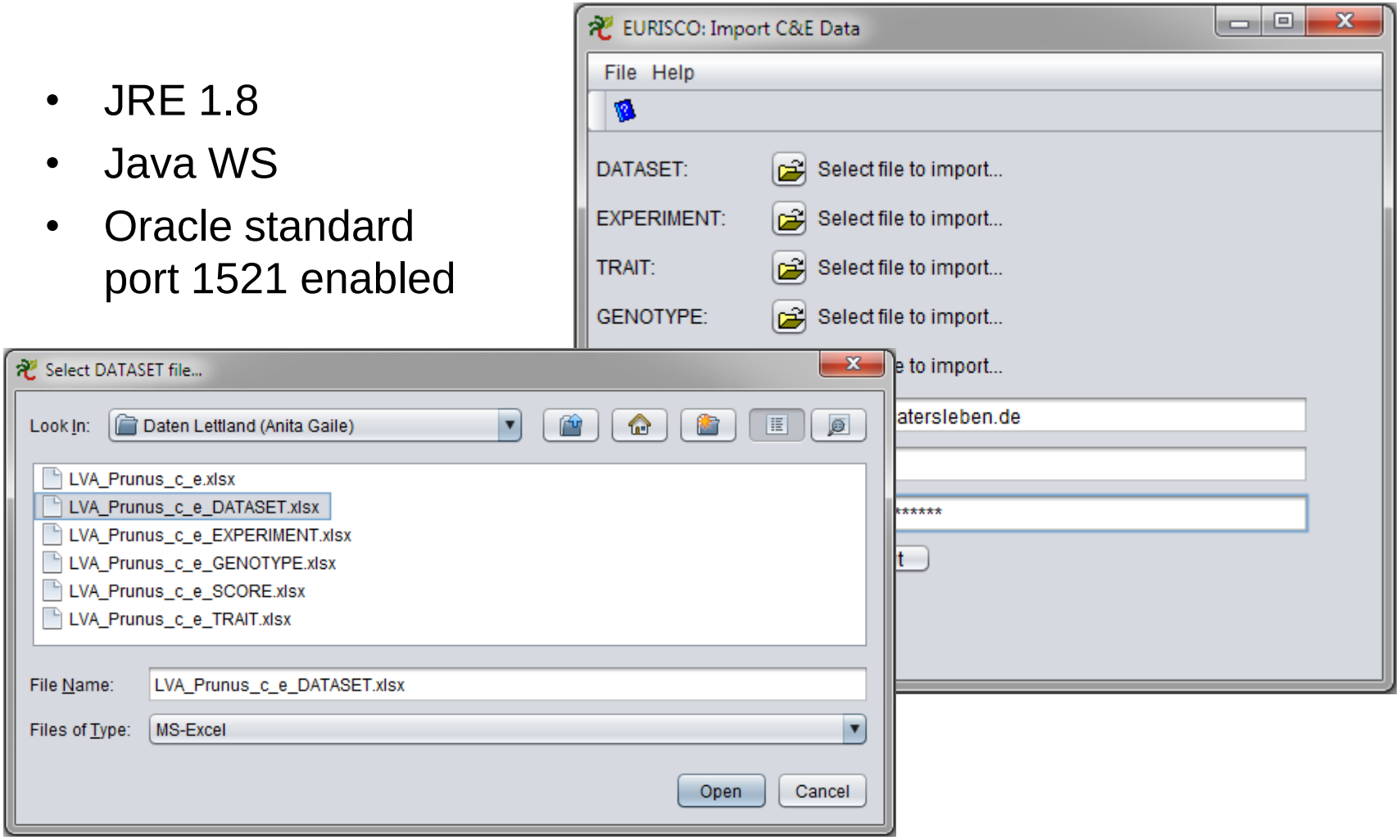


# Upload of C&E data files I

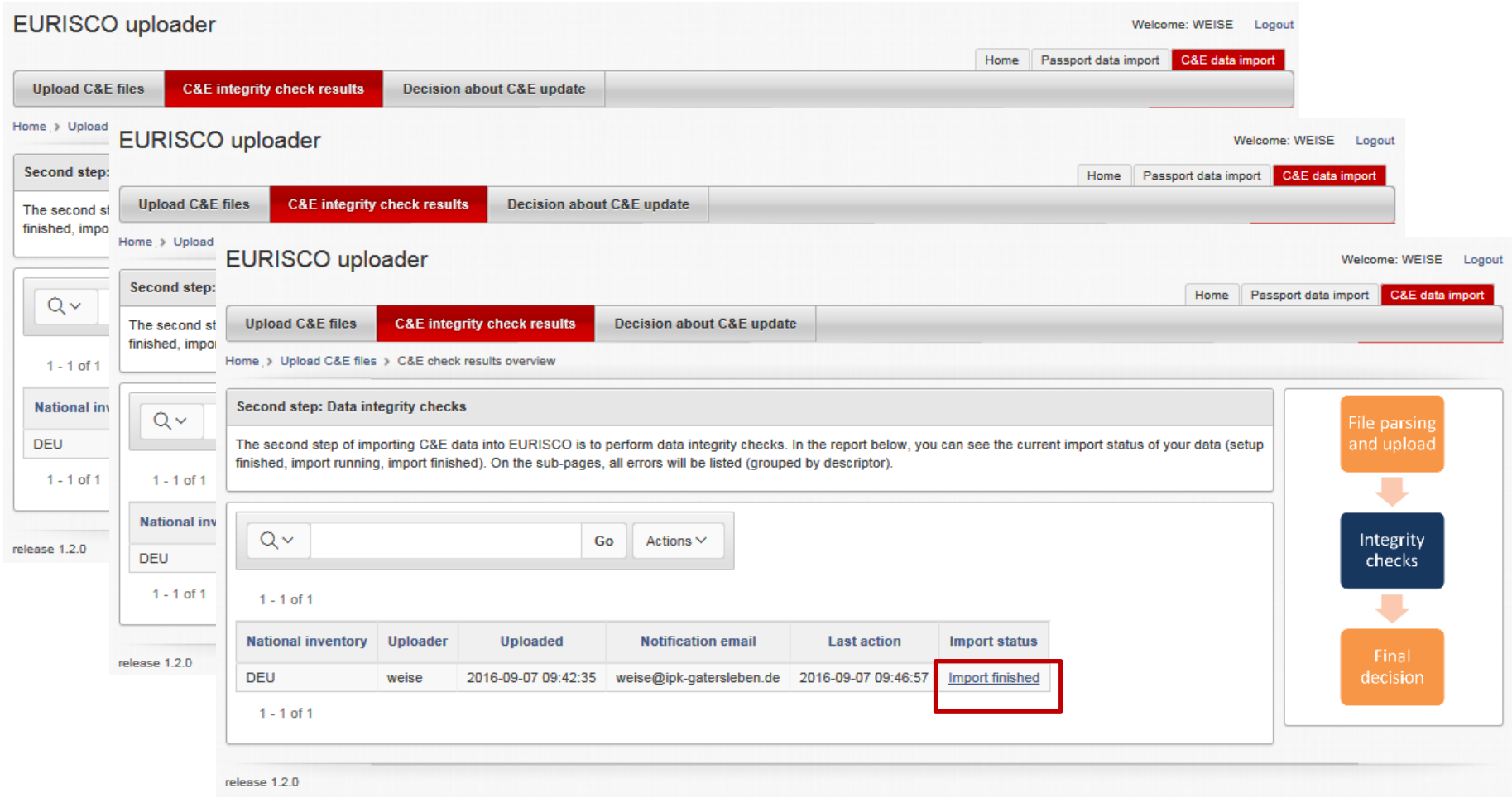


# Upload of C&E data files II

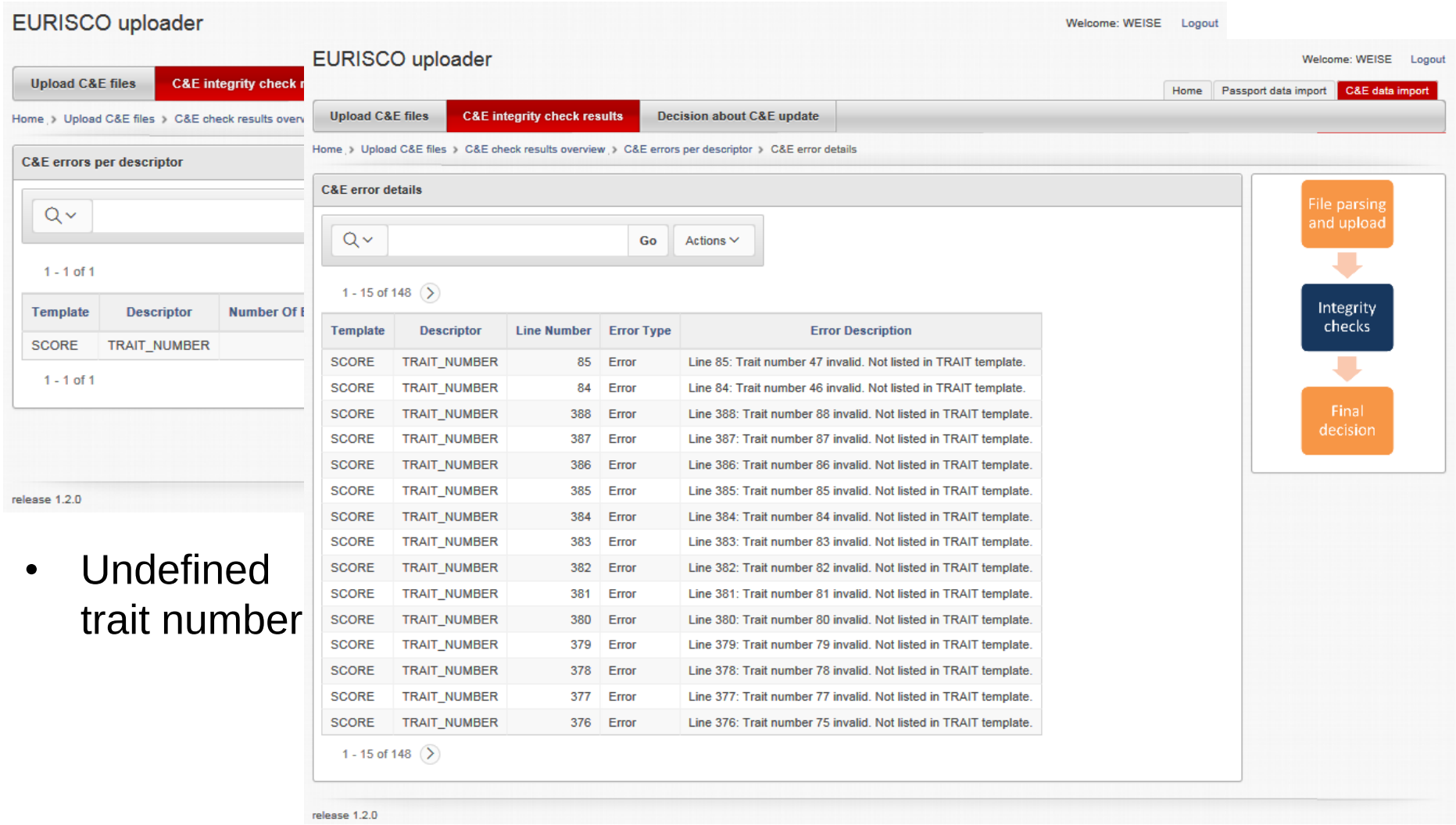
- JRE 1.8
- Java WS
- Oracle standard port 1521 enabled



# Integrity checks I

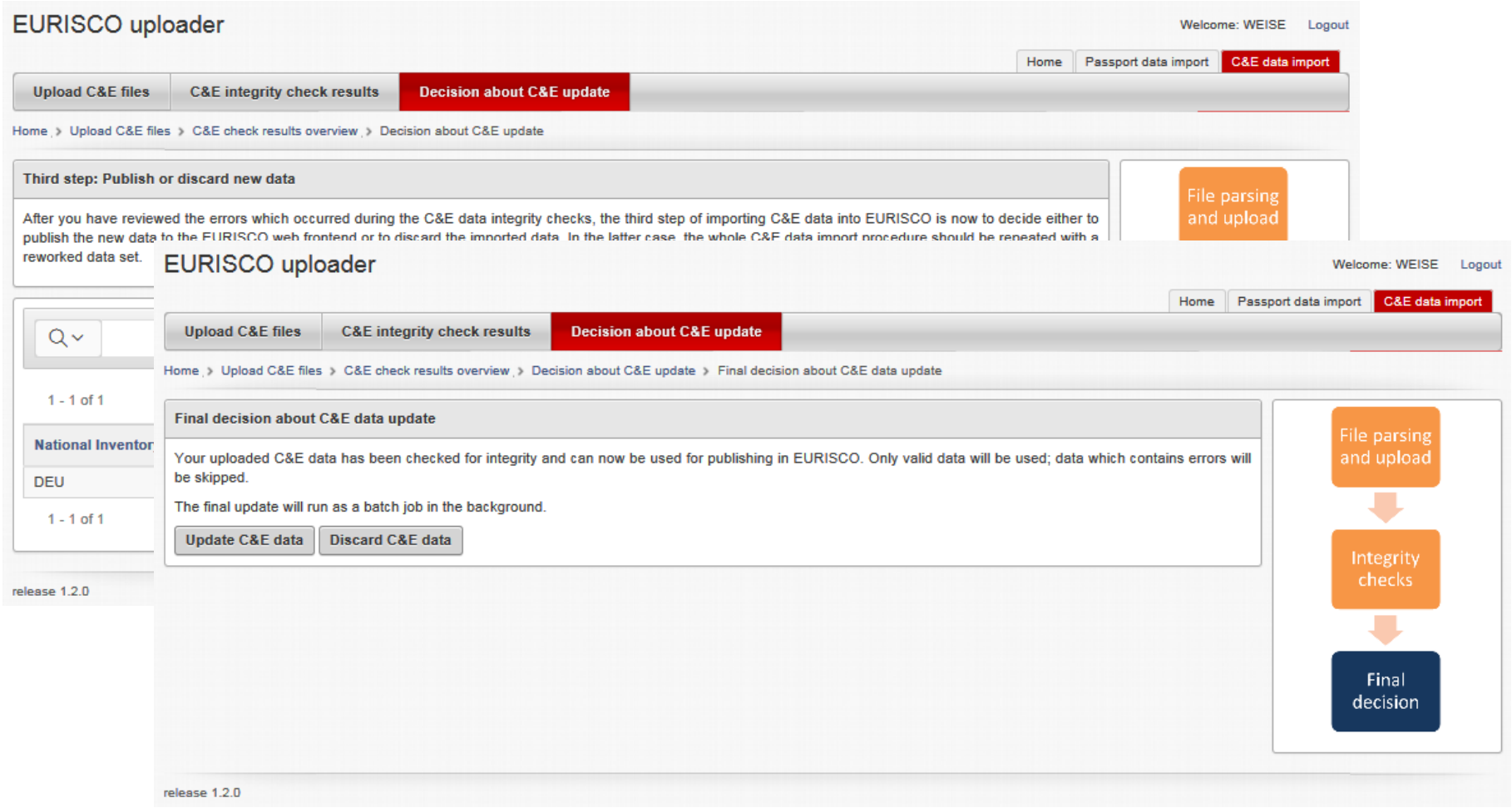


# Integrity checks II




- Undefined trait number

# Final decision



# Next steps (in background)

- New dataset will be applied to EURISCO stage schema
  - Existing C&E data will not be overwritten
  - Existing C&E data may be removed on request

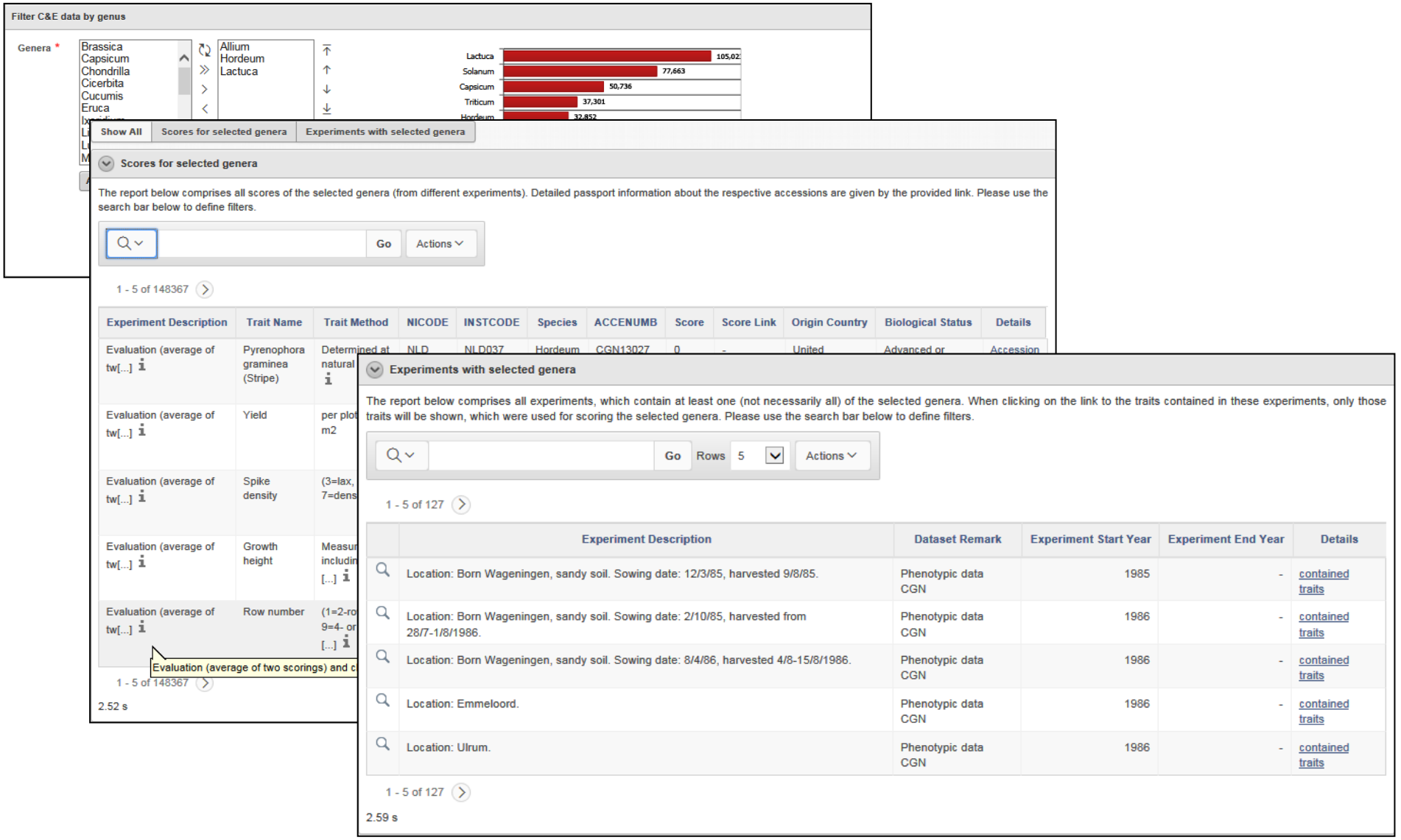
- EURISCO stage will be synchronised to the EURISCO web schema (Time lag!)
  - Not in main business hours
  - Rebuild of materialised views
  - News message on EURISCO webpage

# DATA VISUALISATION

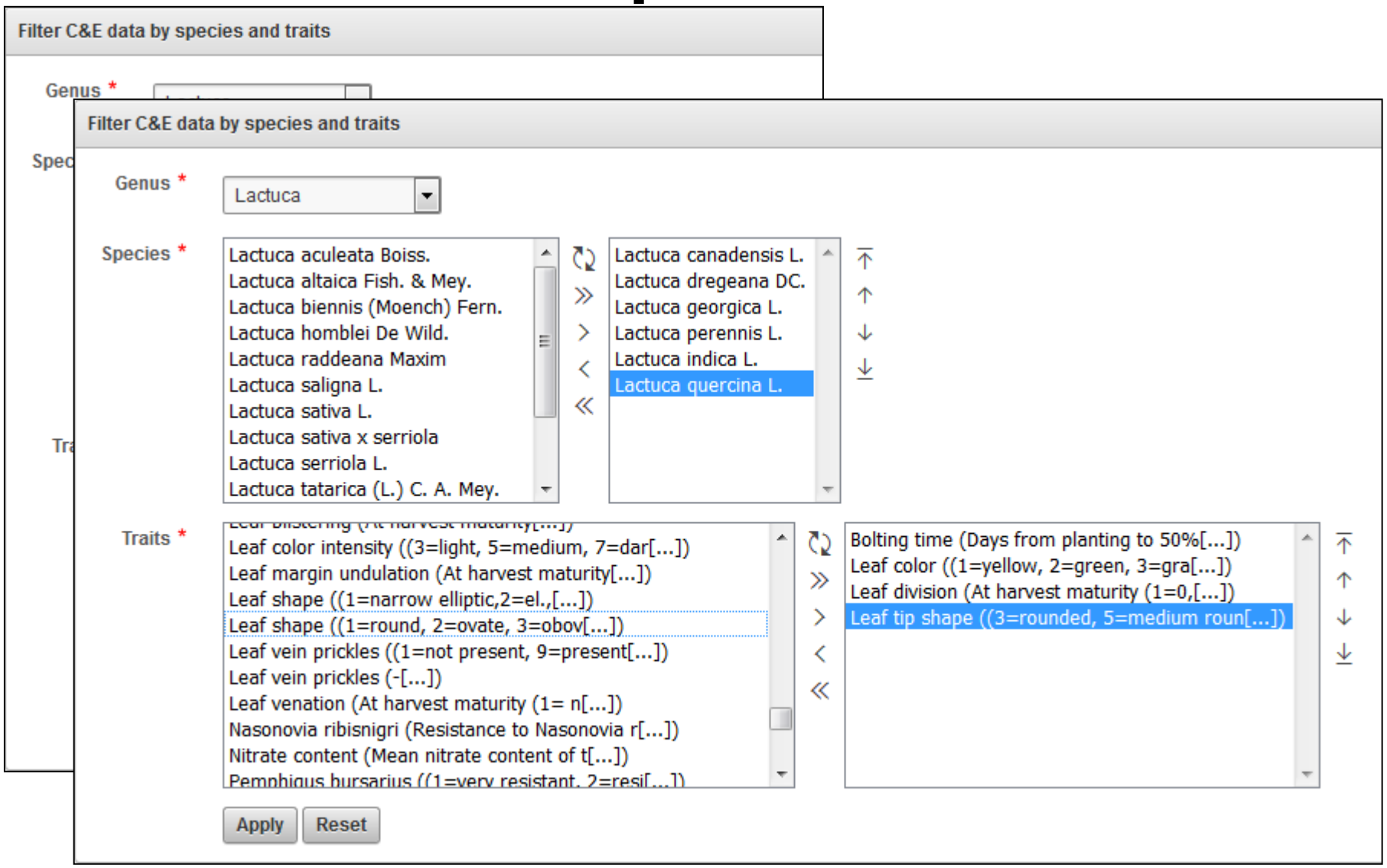
# EURISCO web interface

- Wizard-based searches for C&E data
  - By genus
  - By species and traits
  - By experiment
  - By trait

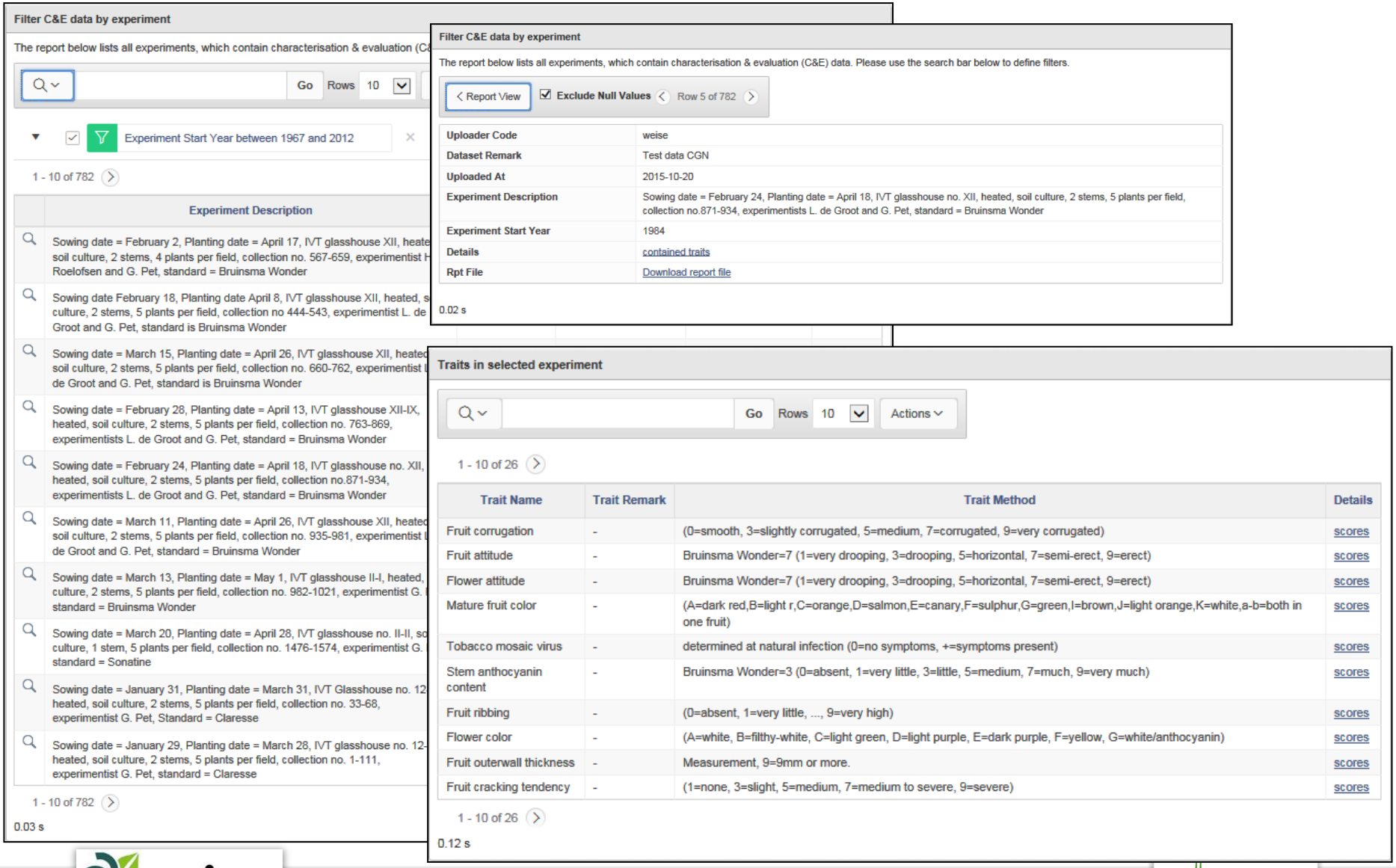
# C&E data: Genus



# C&E data: Species and trait



# C&E data: Experiment



# C&E data: Trait

**Filter C&E data by trait**

The report below lists the definitions of all phenotypic traits, which are currently available in EURISCO. Please use the search bar below to define filters.

Go Rows 10 Actions

1 - 10 of 1214

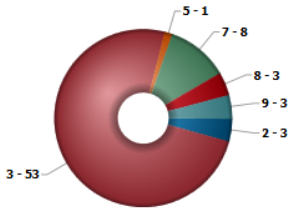
| Trait Name | Trait Remark | Trait Method | Trait Group | Details |
|---|---|---|---|---|
| Petiole and/or midvein enlargement | - | 3=narrow, 5=intermediate,7=enlarged (See IPGRI descriptors Brassica and Raphanus 1990 4.2.27) | C&E data (not further specified) | used by experiment(s) |
| Silique angle | - | 1=0°; 2=0-22.5°; 3=22.5°; 4=22.5-45°; 5=45°; 6:45-67.5°; 7=67.5° | C&E data (not further specified) | used by experiment(s) |
| Leaf anthocyanin content | - | 1 = very weak, 3 = weak, 5 = medium, 7 = strong, 9 = very strong | C&E data (not further specified) | used by experiment(s) |
| Time period until marketable maturity | - | (1=very early, ..., 9=very late) | C&E data (not further specified) | used by experiment(s) |
| Bremia lactucae | - | Resistance to B. lactucae race BI 5, scale:1 = susceptible, 9 = resistant | C&E data (not further specified) | used by experiment(s) |
| Bremia lactucae | - | Resistance to B. lactucae race BI 11, scale:1 = susceptible, 9 = resistant | C&E data (not further specified) | used by experiment(s) |
| Bremia lactucae | - | Resistance to B. lactucae race BI 12, scale:1 = susceptible, 9 = resistant | C&E data (not further specified) | used by experiment(s) |
| Bremia lactucae | - | Resistance to B. lactucae race BI 14, scale:1 = susceptible, 9 = resistant | C&E data (not further specified) | used by experiment(s) |
| Leaf anthocyanin distribution | - | 1 = localised, 2 = entire | C&E data (not further specified) | used by experiment(s) |
| Bremia lactucae | - | Resistance to B. lactucae race BI 25, scale:1 = susceptible, 9 = resistant | C&E data (not further specified) | used by experiment(s) |

1 - 10 of 1214

0.03 s

# C&E data: C&E scores

**Trait details**




**Descriptive statistics**

| Trait Name | Minimum | Maximum | Average | Stddev | Variance | First Quartile | Median | Third Quartile |
|---|---|---|---|---|---|---|---|---|
| Fruit attitude | 2 | 9 | 3.9 | 1.93 | 3.72 | 3 | 3 | 3 |

Experiment description: Sowing date = February 2, Planting date = April 17, IVT glasshouse XII, heated, soil culture, 2 stems, 4 plants per field, collection no. 567-659, experimentist H. Roelofsen and G. Pet, standard = Bruinsma Wonder

Trait name: Fruit attitude

**Additional filters**

Genus: -- All genera of selected trait --

Origin Country: -- All origin countries of selected trait --

**Accession scores for selected trait**

Go | Rows 10 | Actions

1 - 10 of 71

| NICODE | INSTCODE | GENUS | ACCENUMB | Score | Score Link | Origin Country | Biological Status | Details |
|---|---|---|---|---|---|---|---|---|
| NLD | NLD037 | Capsicum | CGN16913 | 3 | - | Germany | Advanced or improved cultivar (conventional breeding methods) | Accession details |
| NLD | NLD037 | Capsicum | CGN16914 | 3 | - | Netherlands | Advanced or improved cultivar (conventional breeding methods) | Accession details |
| NLD | NLD037 | Capsicum | CGN16916 | 8 | - | Israel | Advanced or improved cultivar (conventional breeding methods) | Accession details |
| NLD | NLD037 | Capsicum | CGN16917 | 7 | - | Israel | Advanced or improved cultivar (conventional breeding methods) | Accession details |
| NLD | NLD037 | Capsicum | CGN16918 | 3 | - | Hungary | Traditional cultivar/landrace | Accession details |
| NLD | NLD037 | Capsicum | CGN16919 | 3 | - | Hungary | Advanced or improved cultivar (conventional breeding methods) | Accession details |
| NLD | NLD037 | Capsicum | CGN16920 | 7 | - | - | - | Accession details |
| NLD | NLD037 | Capsicum | CGN16904 | 9 | - | - | - | Accession details |
| NLD | NLD037 | Capsicum | CGN16905 | 3 | - | - | Traditional cultivar/landrace | Accession details |
| NLD | NLD037 | Capsicum | CGN16906 | 3 | - | - | Advanced or improved cultivar (conventional breeding methods) | Accession details |

1 - 10 of 71

0.02 s

# EURISCO web interface

- Still interested in getting additional use-cases for
  - Presentation of C&E data
  - Searching/filtering C&E data
  - Analysis of C&E data

# Examples I

- Wizard-based (step-by-step) selection of C&E data?
  - By taxonomy
  - By trait
  - By experiment
  - …
- Download of selected data (incl. metadata)?
- Descriptive statistics (on experiment level only)?
  - Min, max, avg, stddev, …
  - Box plots, histograms, scatterplots, …
- Comparisons?
  - Different countries of origin
  - …

# Examples II

Go Actions

1 - 15

| Trait Name | Unit | Minimum | Maximum | Average | Stddev | Variance | First Quartile | Median | Third Quartile |
|---|---|---|---|---|---|---|---|---|---|
| Length_primary_root_day_2 | cm | .0769 | 5.0049 | 1.6267 | 1.0375 | 1.0765 | .8328 | 1.3749 | 2.22075 |
| Length_primary_root_day_3 | cm | .145 | 8.2812 | 2.9827 | 1.7086 | 2.9193 | 1.63035 | 2.7457 | 4.010925 |
| Length_primary_root_day_6 | cm | .5243 | 19.8785 | 9.9115 | 3.4562 | 11.9454 | 7.6231 | 9.9545 | 12.1373 |
| Length_primary_root_day_8 | cm | .5973 | 23.1517 | 14.9772 | 4.6948 | 22.0415 | 11.8333 | 15.5338 | 18.4976 |
| Length_primary_root_day_10 | cm | 2.7703 | 24.9071 | 18.5805 | 4.4733 | 20.0107 | 15.7547 | 20.2375 | 21.9946 |
| Length_lateral_roots_day_2 | cm | 0 | .542 | .0036 | .0379 | .0014 | 0 | 0 | 0 |
| Length_lateral_roots_day_3 | cm | 0 | .6299 | .0057 | .0444 | .002 | 0 | 0 | 0 |
| Length_lateral_roots_day_6 | cm | 0 | 42.247 | 4.1527 | 5.8309 | 33.999 | 0 | 1.8464 | 6.096825 |
| Length_lateral_roots_day_8 | cm | 0 | 130.4067 | 28.495 | 24.3938 | 595.056 | 10.6366 | 22.1419 | 37.7196 |
| Length_lateral_roots_day_10 | cm | 2.4513 | 235.7843 | 66.2073 | 51.863 | 2689.7719 | 27.4616 | 50.745 | 88.5509 |
| Total_root_length_day_2 | cm | .0769 | 5.0049 | 1.6307 | 1.0403 | 1.0822 | .8328 | 1.3749 | 2.22995 |
| Total_root_length_day_3 | cm | .145 | 8.2812 | 2.9884 | 1.7113 | 2.9286 | 1.636425 | 2.7457 | 4.027575 |
| Total_root_length_day_6 | cm | .5243 | 59.3962 | 14.0647 | 8.3458 | 69.652 | 8.49165 | 11.8062 | 17.34515 |
| Total_root_length_day_8 | cm | 5.0457 | 151.2535 | 43.4908 | 27.7073 | 767.693 | 22.3352 | 37.1098 | 55.4076 |
| Total_root_length_day_10 | cm | 10.5623 | 268.6712 | 85.2913 | 55.3723 | 3066.0905 | 43.067 | 69.7821 | 110.9772 |

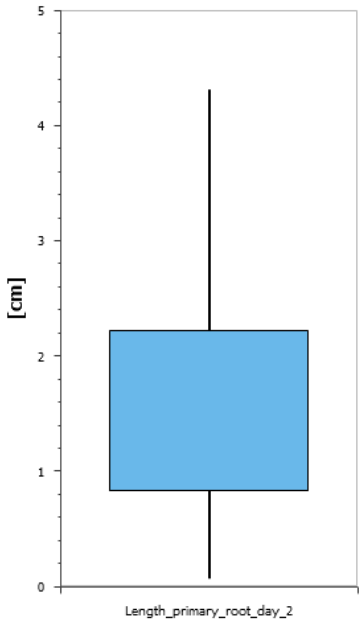
1 - 15

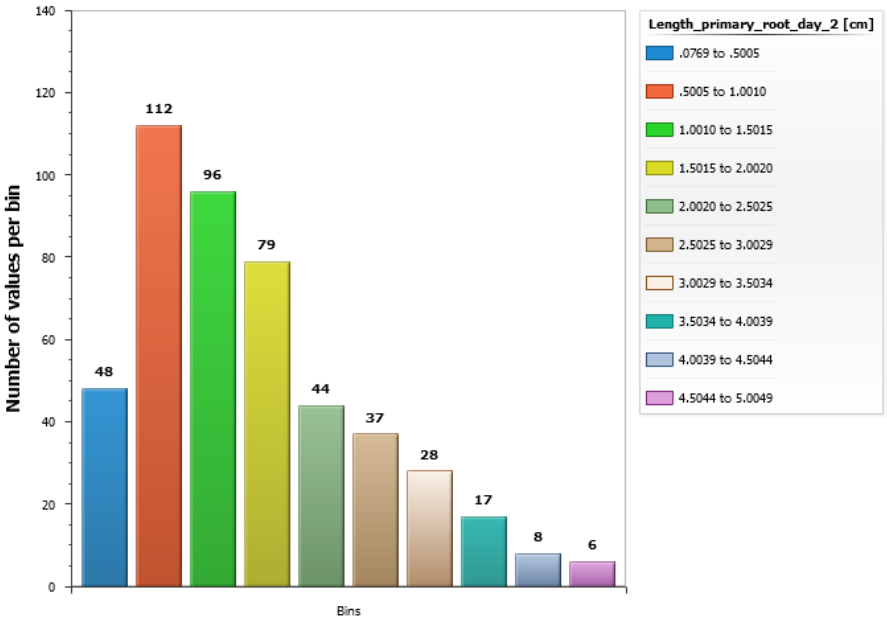
0.30 s

# Examples III

Select trait: Length_primary_root_day_2

| Trait | Unit | Minimum | Maximum | Average | Stddev | Variance | First Quartile | Median | Third Quartile |
|---|---|---|---|---|---|---|---|---|---|
| Length_primary_root_day_2 | cm | .0769 | 5.0049 | 1.6267 | 1.0375 | 1.0765 | .8328 | 1.3749 | 2.22075 |

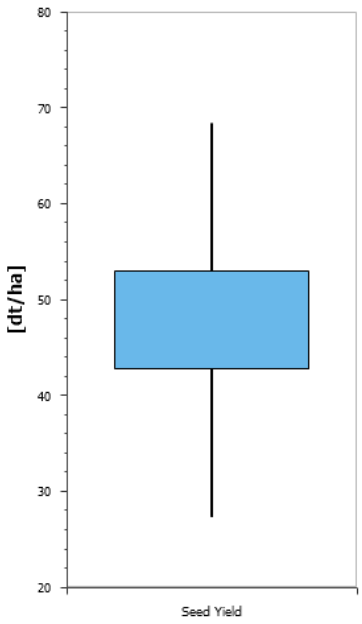


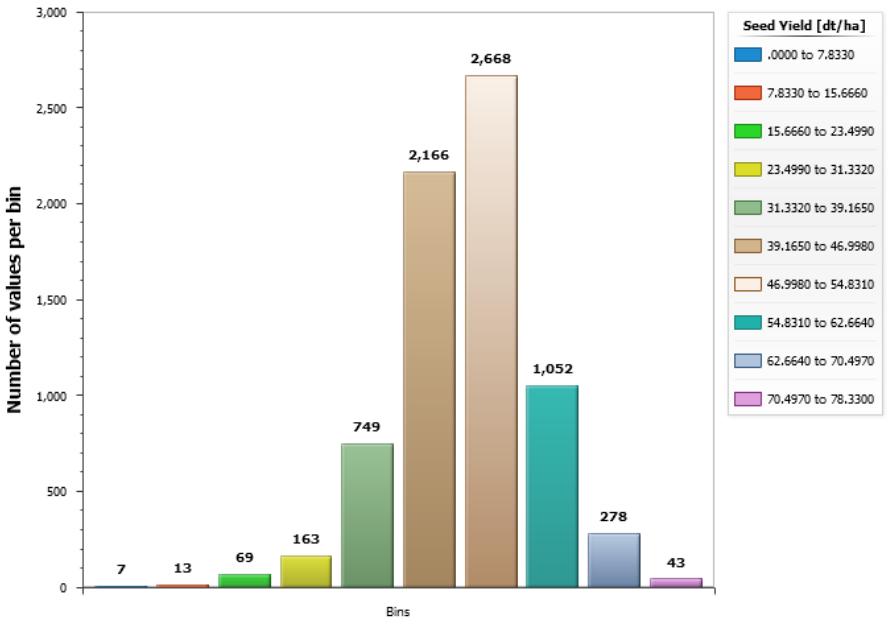



# Examples IV

Select trait: Seed Yield

| Trait | Unit | Minimum | Maximum | Average | Stddev | Variance | First Quartile | Median | Third Quartile |
|---|---|---|---|---|---|---|---|---|---|
| Seed Yield | dt/ha | 0 | 78.33 | 47.8293 | 8.854 | 78.3934 | 42.76725 | 48.177 | 52.983975 |

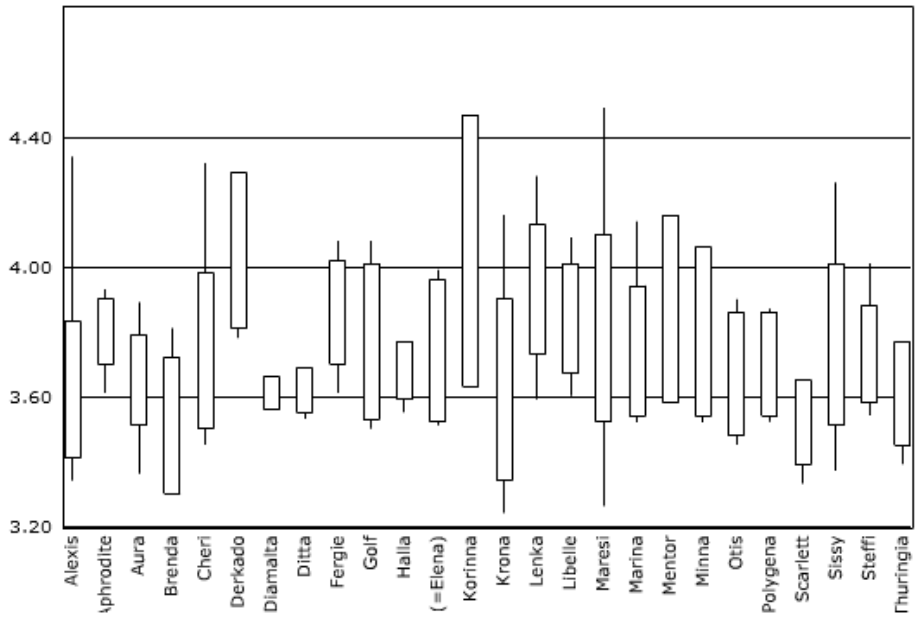


# Examples V

Averages per trait

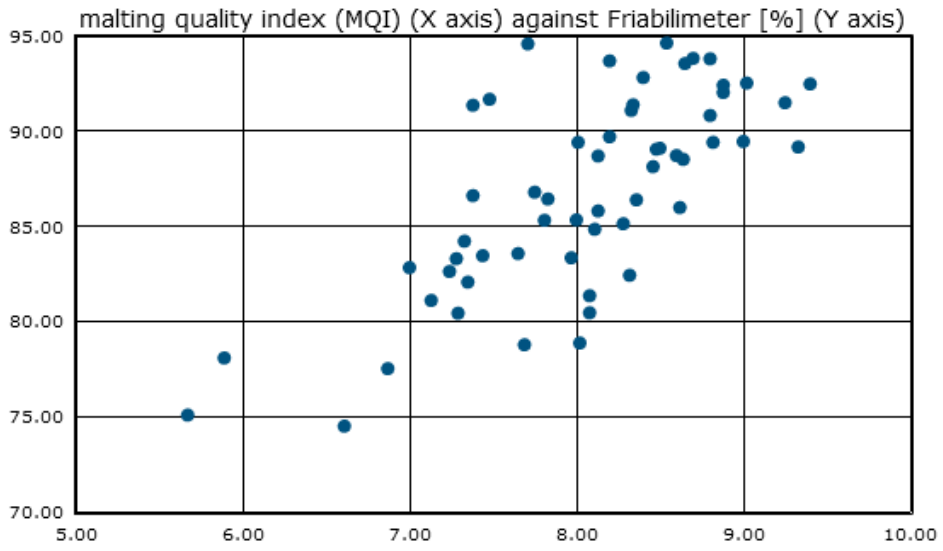
real extract [%] | 2-rowed | spring barley | Plot



# Examples VI

Scatterplot

malting quality index (MQI) | Friabilimeter [%] | 2-rowed | spring barley | Plot



# What are your requirements?