Integration of C&E data into EURISCO
## (including IDBB data)

BETANET Meeting, 19–20 June 2018, Venice, Italy

Stephan Weise

Background

- What is EURISCO?
  - European information system for plant genetic resources
  - Search catalogue for *ex situ* collections
  - Accession-level information system

- Purpose
  - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe
  - Assists in meeting national obligations
    - Food and Agriculture Organization of the United Nations (FAO)
    - Convention on Biological Diversity (CBD)
    - International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA)
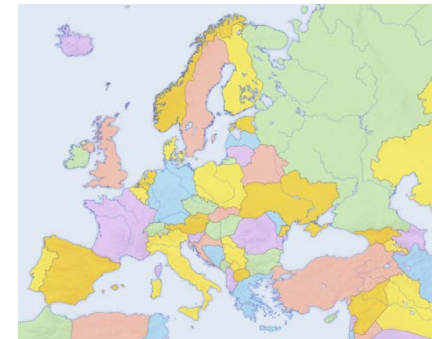
https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png

Data flow
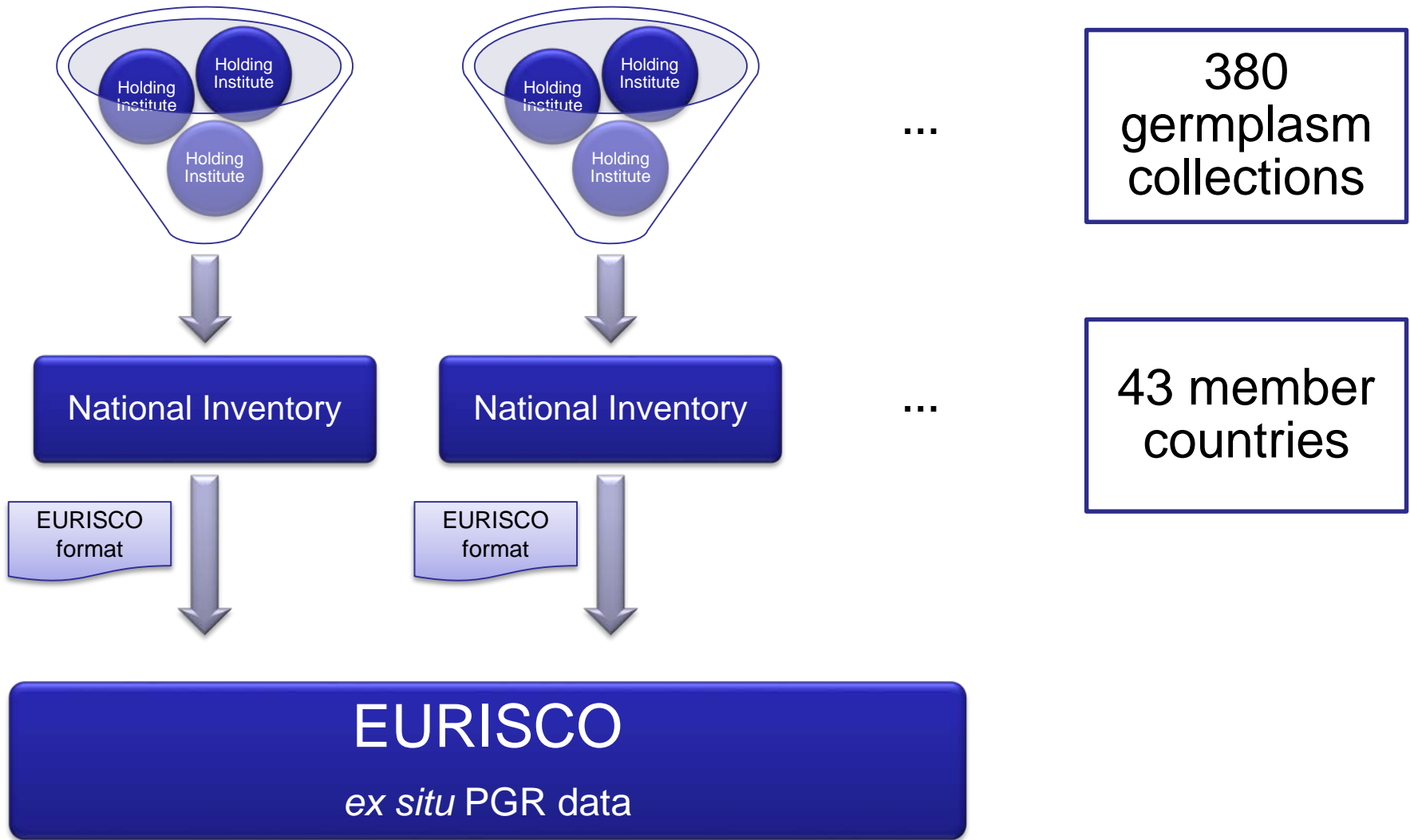
Contents of EURISCO

- 1,973,603 accessions

- 6,438 genera
  (including synonyms, spelling variants)

- 43,994 species names
  (unique combinations genus + species, including synonyms)

- 419,789 MLS accessions

- 34,364 AEGIS accessions

- 22,906 DOIs

as at 2018-06-06

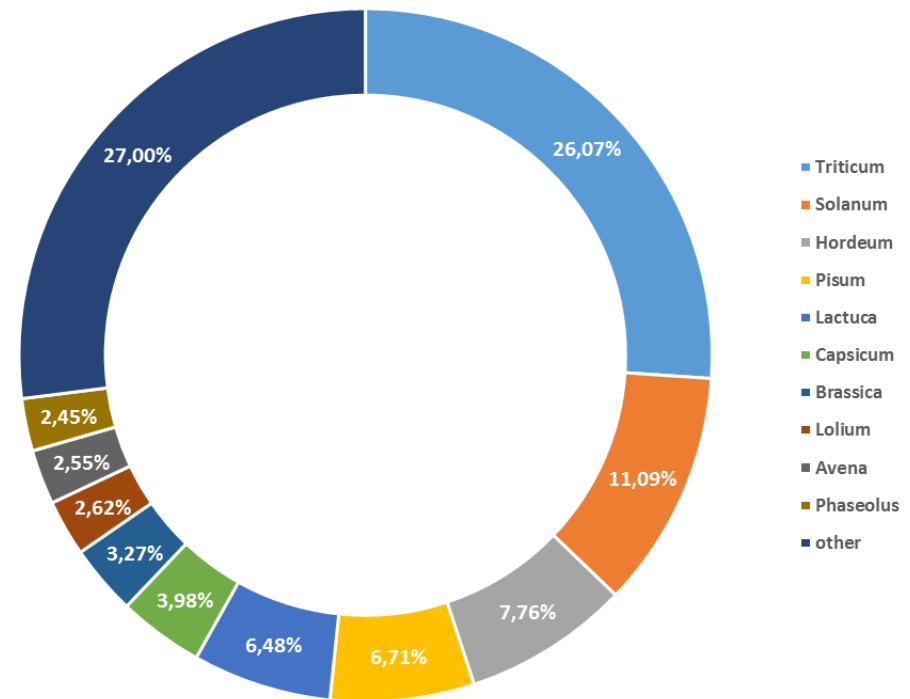*Weise et al. (2017) Nucleic Acids Research, 45(D1):D1003-D1008.*

LEIBNIZ INSTITUTE OF PLANT GENETICS AND CROP PLANT RESEARCH

Passport data in EURISCO

- Four standard searches:

  – Taxonomy

  – Accession

  – Biological status

  – Collecting site

- Advanced search

- Different user-specific export features

C&E data in EURISCO

- Currently, 1,652,895 records of data from seven countries
  - Czech Republic
  - Germany
  - Latvia
  - The Netherlands
  - Poland
  - Romania
  - United Kingdom

- 68,821 accs. with C&E data

Legend:
- Triticum — 26,07%
- Solanum — 11,09%
- Hordeum — 7,76%
- Pisum — 6,71%
- Lactuca — 6,48%
- Capsicum — 3,98%
- Brassica — 3,27%
- Lolium — 2,62%
- Avena — 2,55%
- Phaseolus — 2,45%
- other — 27,00%

as at 2018-06-06

Search for C&E data I

Wizard-based searches for
- Genus
- Species and trait
- Experiment
- Trait

Search for C&E data II

Well known challenges

- 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.

Current EURISCO Approach

- Result of discussions of previous years

- Suitable proposal compiled within the ECPGR Doc&Info working group

- Pragmatic approach: Import of existing data as-is to reach critical mass

  – No standardisation of trait, scale or experimental design

  – Only standardisation of exchange format

    - As simple as possible

    - As few fields as possible

  → "minimum consensus"

Data model for C&E data

GENOTYPE

• Identified by EURISCO descriptors

DATASET

• May comprise different experiments

EXPERIMENT

• Multiple genotypes are scored for different traits

TRAIT

• Characteristic feature to be scored

SCORES

• Value of a trait for a genotype

eurisco — Finding seeds for the future

IPK GATERSLEBEN

Proceeding for data upload

- 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 exchange format in MS Excel (.xlsx) files

- Upload via EURISCO intranet
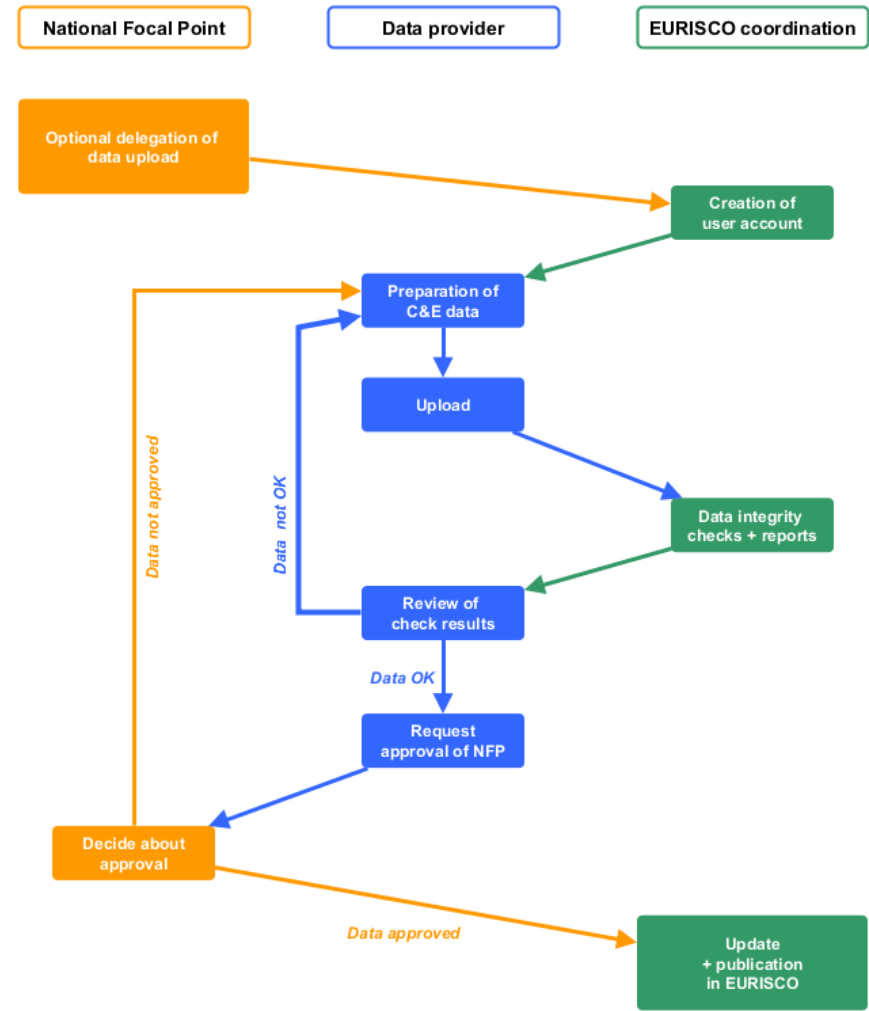
Data upload in three steps

File parsing and upload via Java tool
- data owner

Data integrity checks
- EURISCO management

Approval / withdrawal of data for publishing on the EURISCO website
- *data owner/NFP*

LEIBNIZ INSTITUTE OF PLANT GENETICS AND CROP PLANT RESEARCH
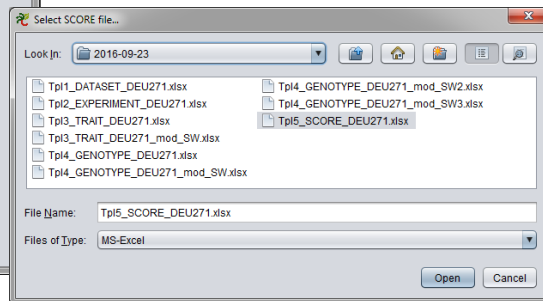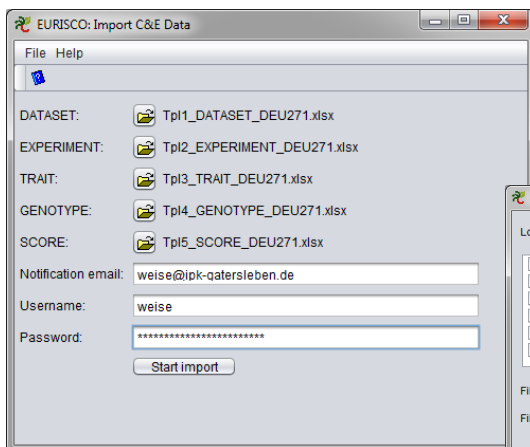
Overview IDBB

- International Data Base for Beet

- Maintained at JKI on behalf of ECPGR's Beta Working Group

- Content:
  - 10,528 accessions of different categories
  - 44,750 observations of varying quality of which about
  - 20,000 observations refer to the GENRES CT95-42 project
    - 689 evaluated accessions
    - 44 descriptors
    - 50 methods
    - 54 experiments

- Elaborated data models published in Beta WG reports (1999, 2002)

Transfer of IDBB C&E data

Transfer of US material to GRIN
- C&E data restricted to accessions documented in EURISCO
- Responsible: L. Frese

Check, which European IDBB accessions are not listed in EURISCO
- Responsible: EURISCO coordination

Check, why accessions are not listed
- Changed identifiers (INSTCODE, GENUS, ACCENUMB)?
- Accs. not existing anymore?
- Other reasons?
- Responsible: NFPs/Beta WG

Import of C&E data into EURISCO
- Responsible: EURISCO coordination
- Final approval by NFPs
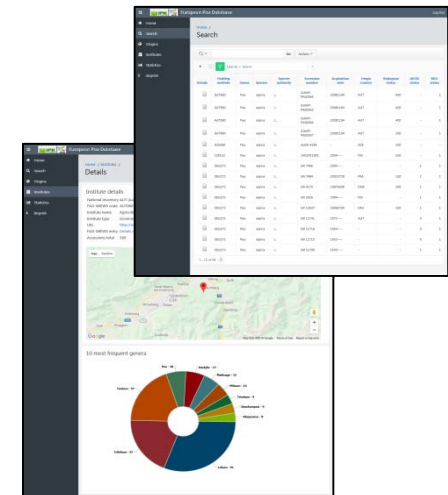
Provision of IDBB C&E data in EURISCO format
- Responsible: C. Germeier

Decide about adding those accessions to EURISCO
- Responsible: NFPs

eurisco
Finding seeds for the future

IPK
GATERSLEBEN

Future of IDBB

- To be decided by the Beta Working Group
- Possible alternatives:
  1. Only usage of search and filter possibilities of EURISCO

     - No additional work
     - Not crop-specific

  2. Beta portal based on EURISCO data
     - Both passport and C&E data provided by API
     - No need to compile Beta data separately
     - Always up-to-date
     - Beta WG can focus on crop-specific information
     - Example: European Poa Database
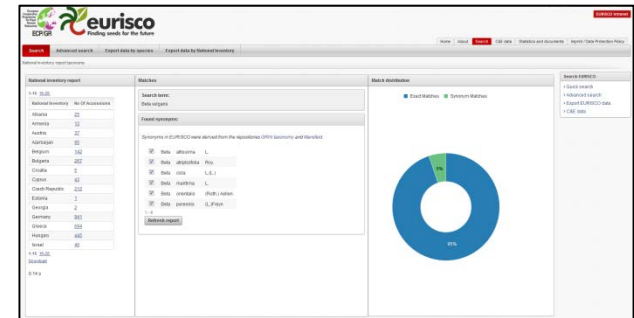
THANK YOU FOR YOUR ATTENTION