EURISCO:
ensuring integration of data in a special intranet environment for EVA

Annual meetings of the European Evaluation Networks (EVA)
2020, online via MS Teams

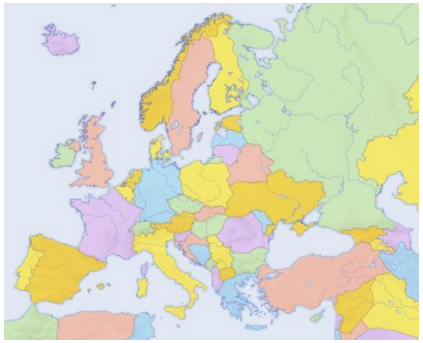
Stephan Weise
30 March 2020

OVERVIEW

Background

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

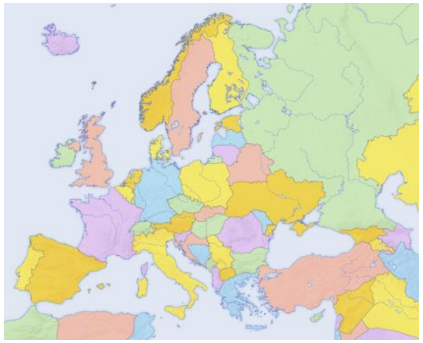
- Purpose
  - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe
  - Assists in meeting national obligations
    - Food and Agriculture Organization of the United Nations (FAO)
    - Convention on Biological Diversity (CBD)
    - International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA)



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

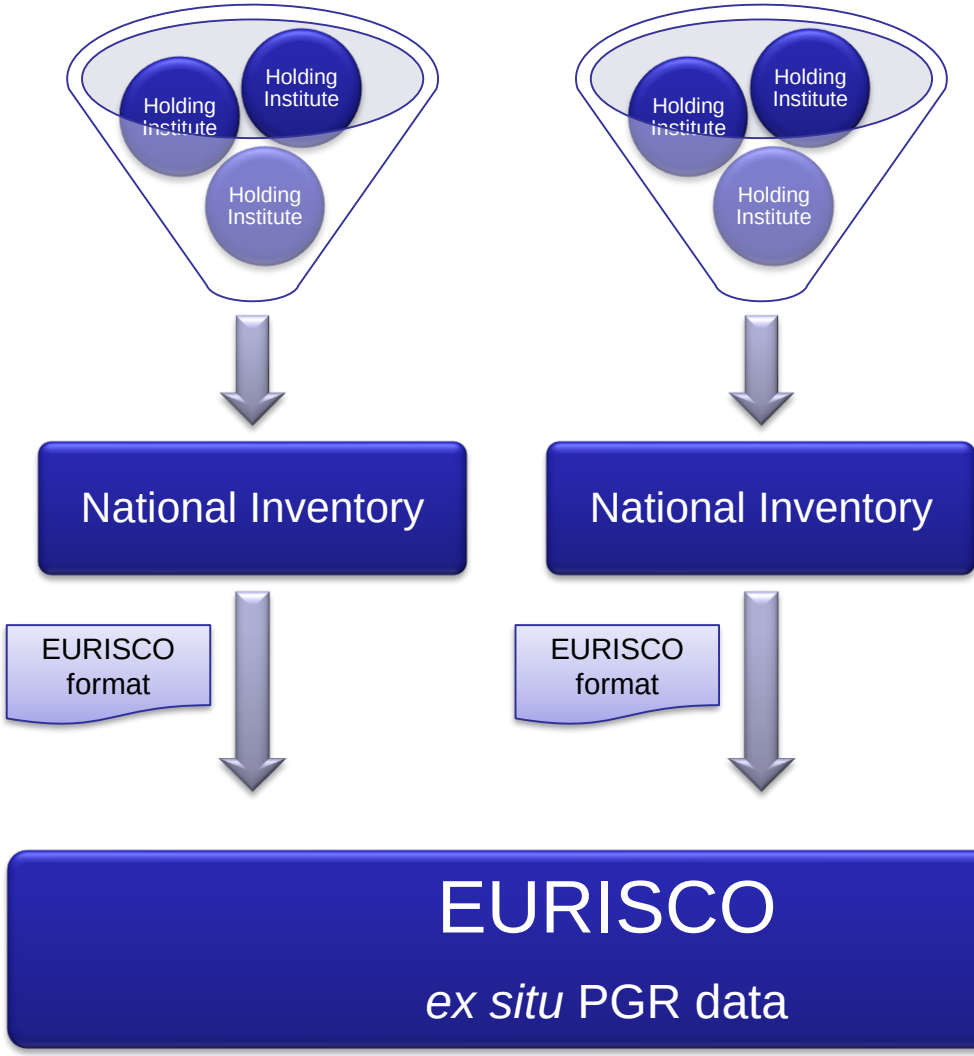
Development

- Started in 1999 (EU project EPGRIS)
- 43 countries involved
  (Nordic Countries → NordGen)
- National collections represented by
  National Inventories (NIs)
- Network of National Focal Points (NFPs)
  links NIs ↔ EURISCO



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

Data flow

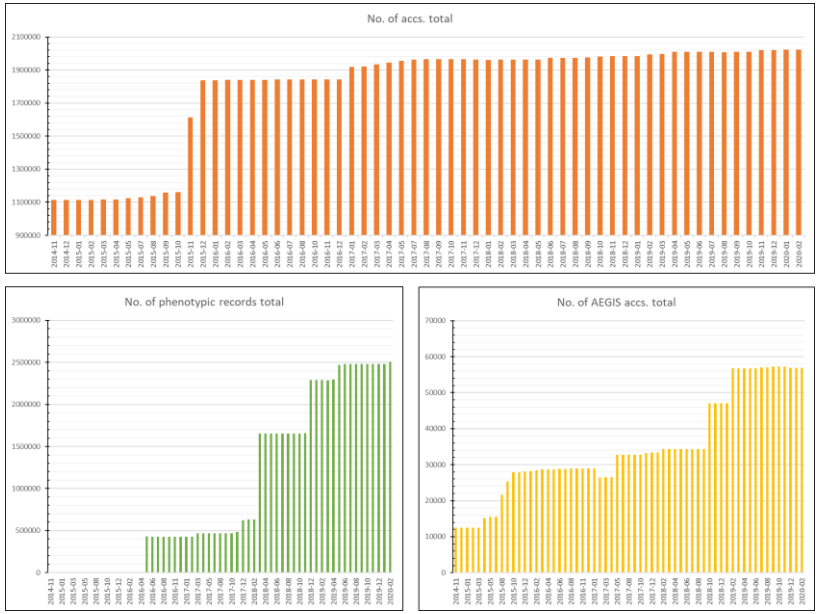


400 germplasm collections

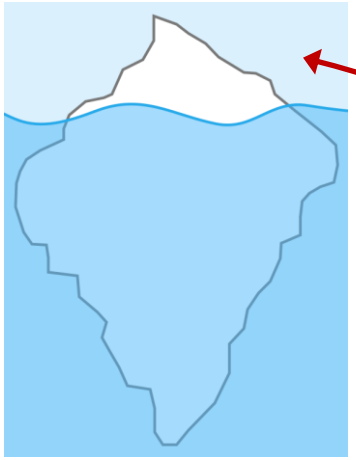
43 member countries

Contents of EURISCO

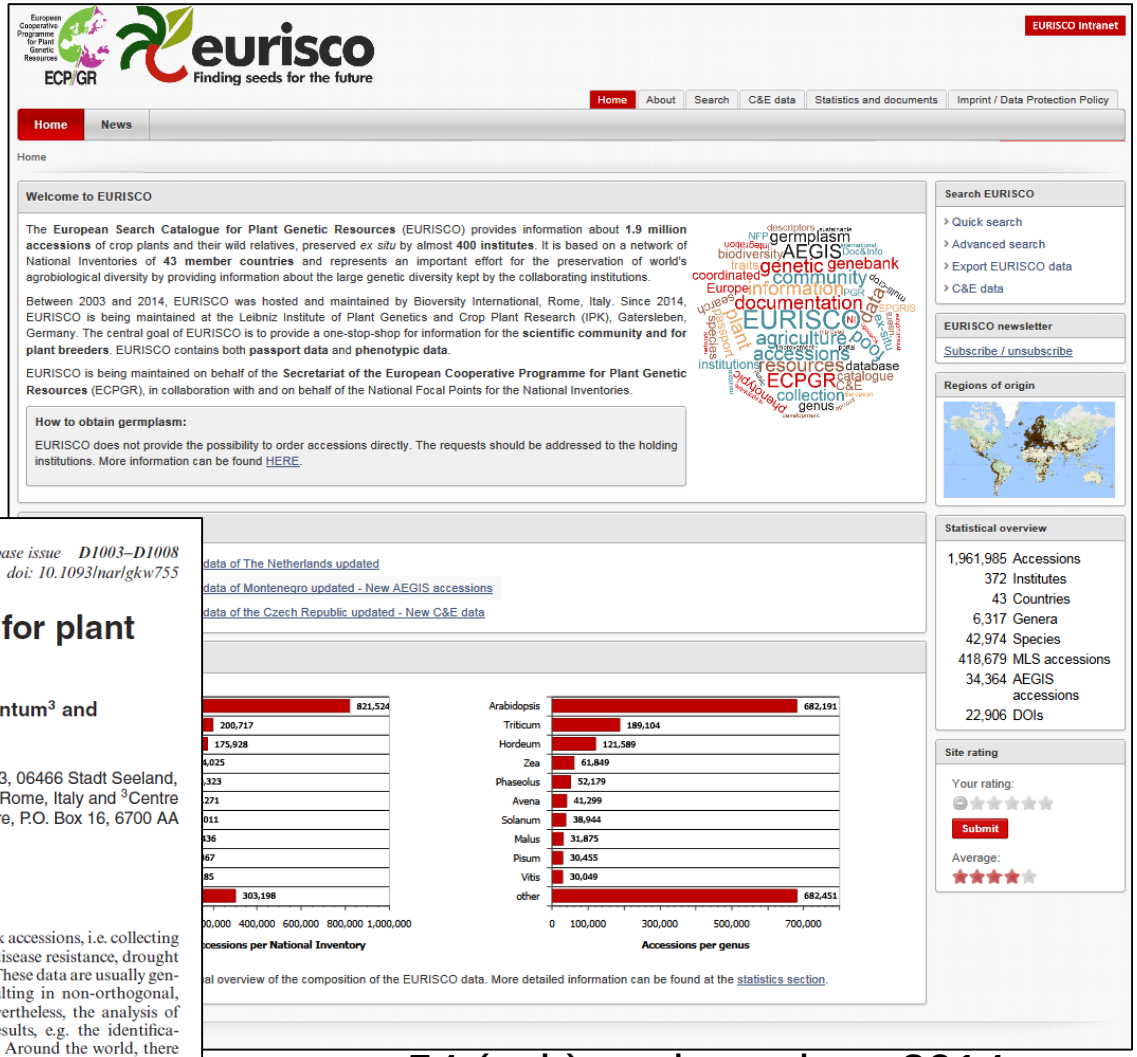
- 2,023,530 accessions
- 6,393 genera
  (including synonyms, spelling variants)
- 43,230 species names
- 443,512 MLS accessions
- 56,928 AEGIS accessions
- 60,500 DOIs



as of 2020-03-25



Web interface



Published online 31 August 2016 *Nucleic Acids Research, 2017, Vol. 45, Database issue* ***D1003–D1008***
doi: 10.1093/nar/gkw755

EURISCO: The European search catalogue for plant genetic resources

Stephan Weise[1,*], Markus Oppermann[1], Lorenzo Maggioni[2], Theo van Hintum[3] and Helmut Knüpffer[1]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Stadt Seeland, Germany, [2]Bioversity International, Via dei Tre Denari 472/a, 00057 Maccarese (Fiumicino), Rome, Italy and [3]Centre for Genetic Resources, The Netherlands (CGN), Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands

Received July 28, 2016; Revised August 19, 2016; Accepted August 21, 2016

ABSTRACT

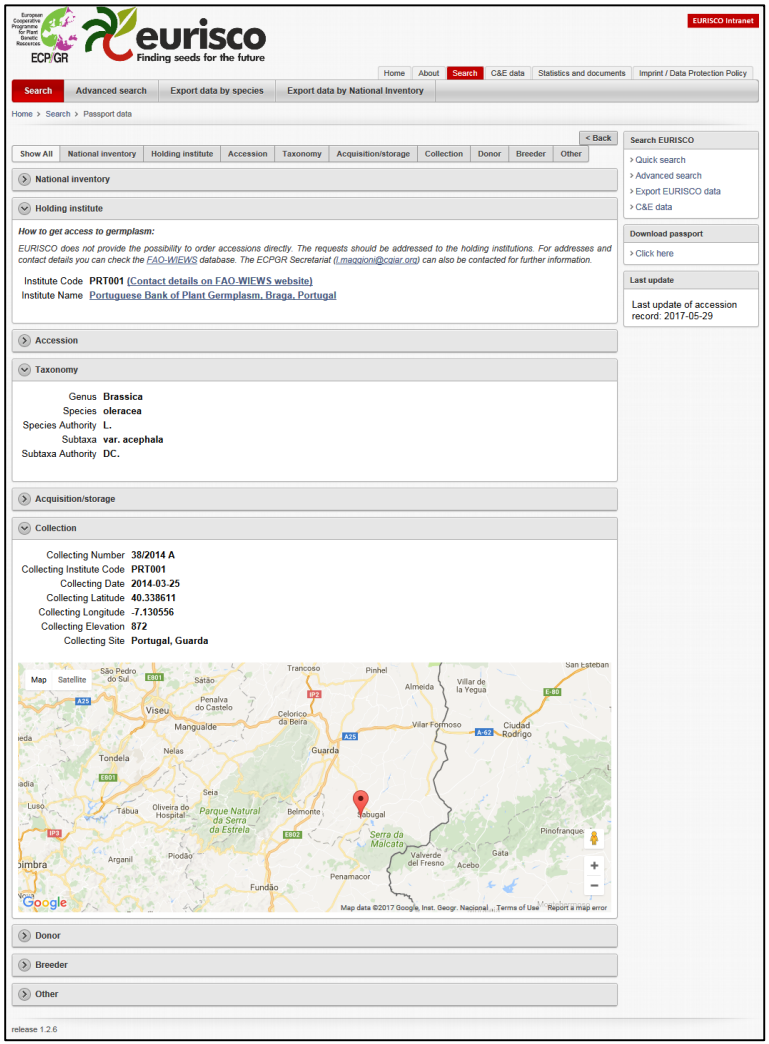
The European Search Catalogue for Plant Genetic Resources, EURISCO, provides information about 1.8 million crop plant accessions preserved by almost 400 institutes in Europe and beyond. EURISCO is being maintained on behalf of the European Cooperative Programme for Plant Genetic Resources. It is based on a network of National Inventories of 43

typic characterisation of genebank accessions, i.e. collecting information about traits such as disease resistance, drought tolerance and yield components. These data are usually generated on selected material, resulting in non-orthogonal, highly incomplete data sets. Nevertheless, the analysis of these data allows meaningful results, e.g. the identification of promising new alleles (5). Around the world, there are about 1800 genebank collections conserving PGRFA. Thereof, about 625 collections are maintained in Europe

54 (sub)versions since 2014

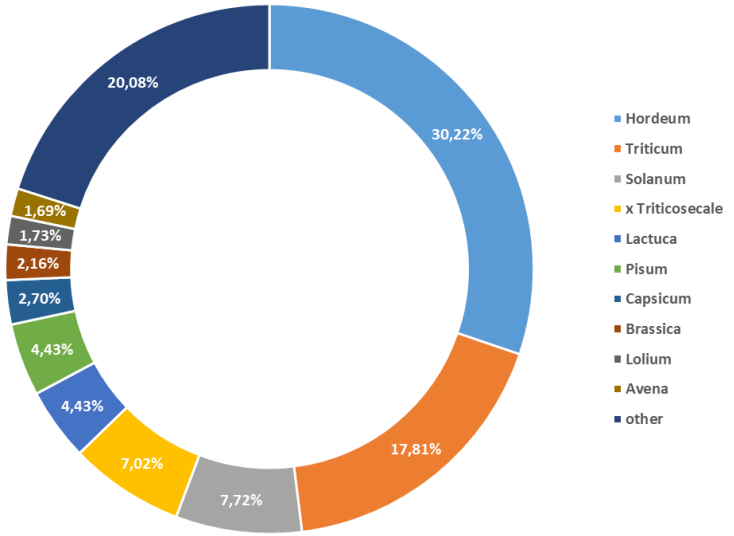
Passport data search in EURISCO

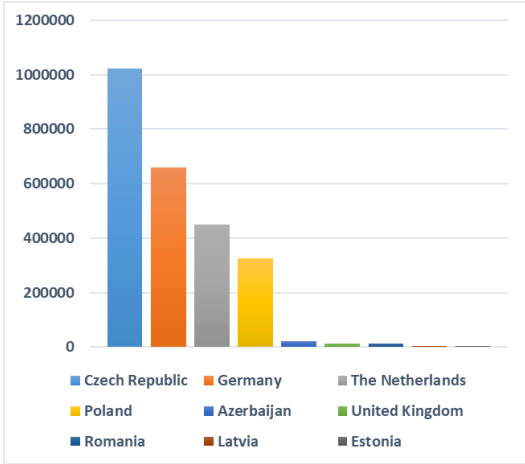
- Four standard searches:
  - Taxonomy (incl. synonyms)
  - Accession
  - Biological status
  - Collecting site
- Advanced search
- Different user-specific export features



Phenotypic data

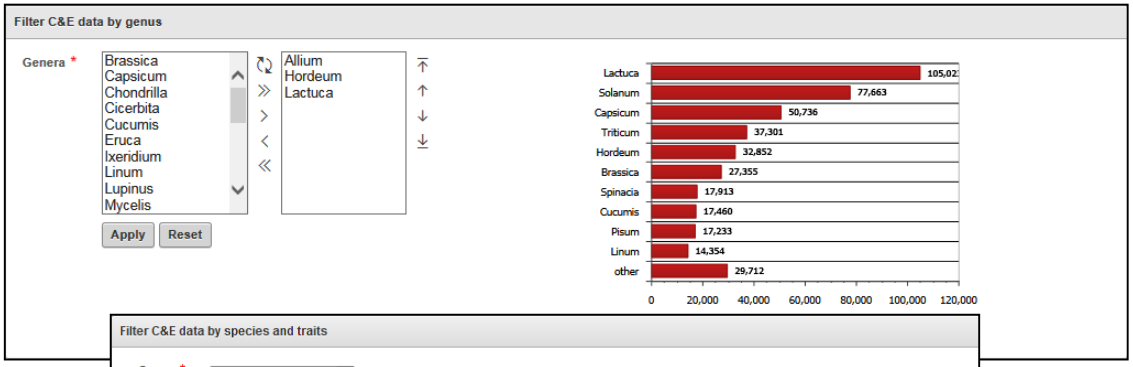
- Extension available since 2016
- Currently, 2,503,655 records of data from nine countries
  - Azerbaijan
  - Czech Republic
  - Estonia
  - Germany
  - Latvia
  - The Netherlands
  - Poland
  - Romania
  - United Kingdom
- 86,507 accs. with phenotypic data

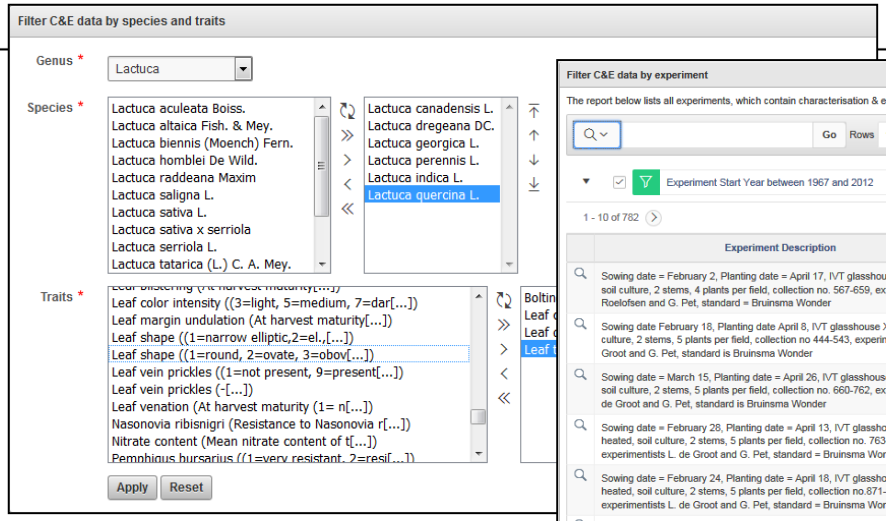


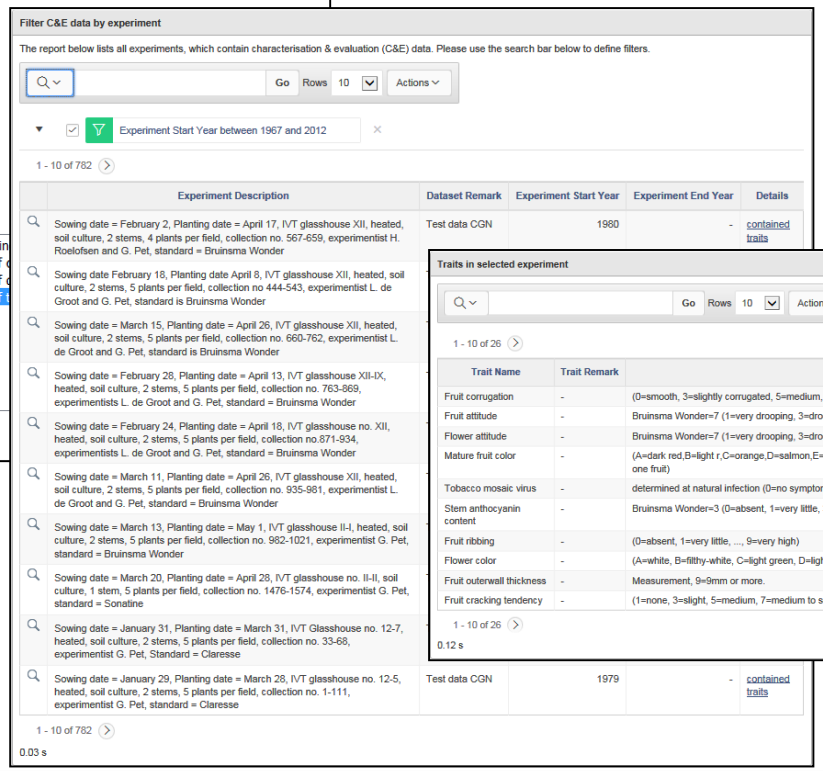


as of 2020-03-26

Phenotypic data search in EURISCO



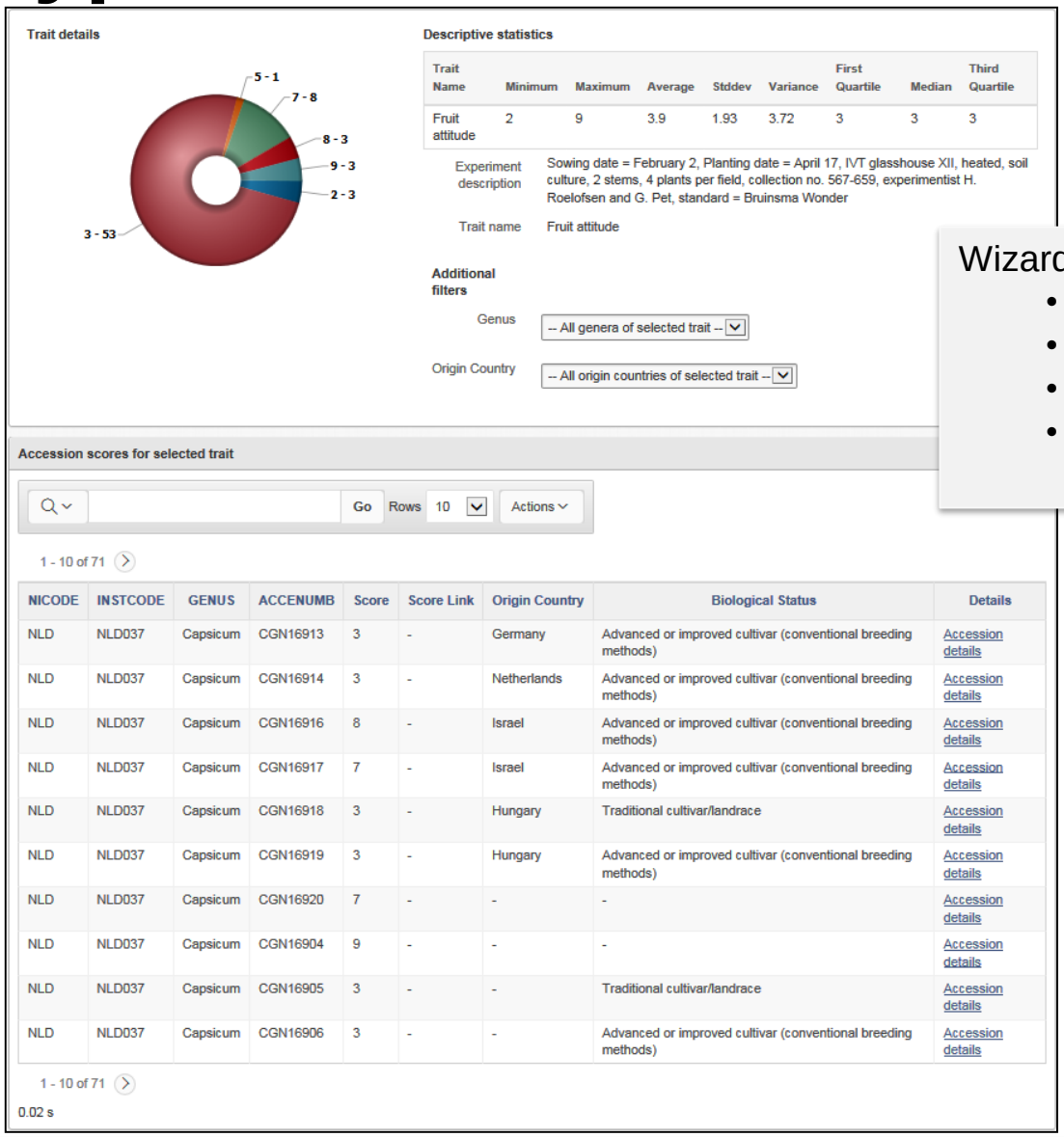




Wizard-based searches for

- Genus
- Species and trait
- Experiment
- Trait

Phenotypic data search in EURISCO



Wizard-based searches for

- Genus
- Species and trait
- Experiment
- Trait

Refine result

- Sort
- Filter
- Download
- Chart

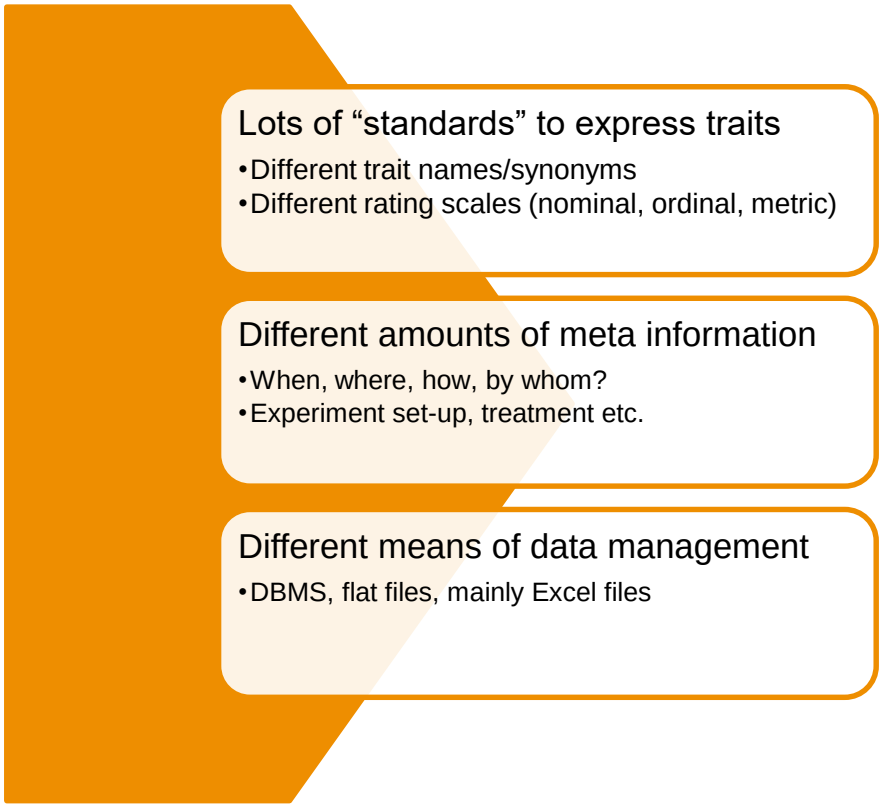
CURRENT LIMITATIONS

Current technical limitations

- Only non-confidential phenotypic data
- Only data of accessions listed in EURISCO
- NFPs must approve data before publication
- No embargo periods

→ Can be solved technically by the EVA intranet

The major challenge: Diversity of data



Existing situation

Methods and Descriptors

- Crop-specific definitions of traits, methods etc. like IPGRI descriptor lists
- Often used in parts only and adapted to organisational needs

Exchange Formats

- E.g. Darwin Core germplasm extension (DwC-germplasm; Endresen et al. 2009)
- Great for computer scientists
- Difficult to handle for genebank curators

Ontologies

- Help to structure the (phenotypic) world
- Improve interoperability of data
- e.g. Crop Ontology (Arnaud et al. 2012)

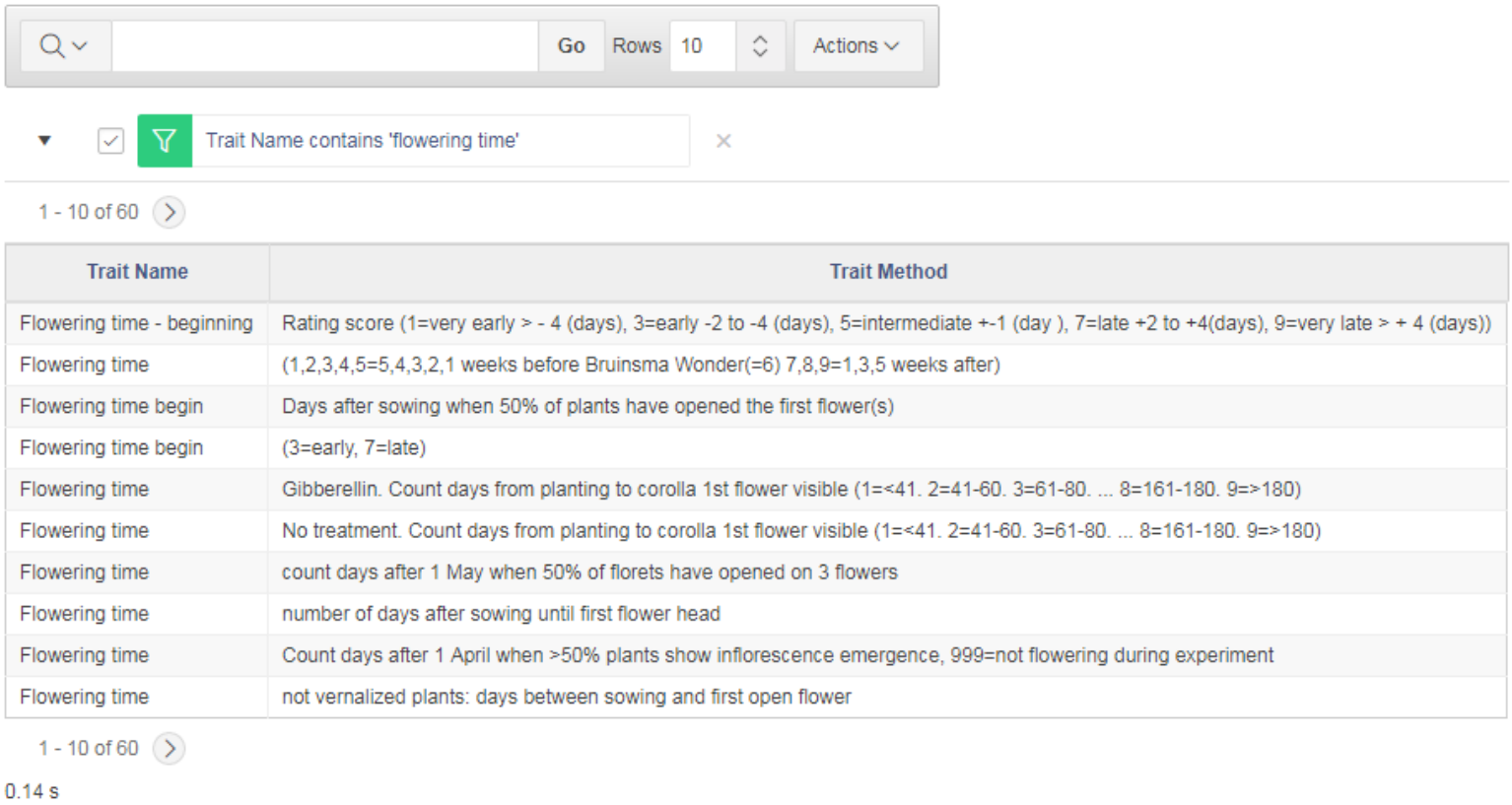
Current EURISCO approach

- Data standardisation
  - No standardisation of trait, scale or experimental design
  - Pragmatic approach: Import of existing data as-is to reach critical mass
- Data exchange
  - Only standardisation of exchange format
    - As simple as possible
    - As few fields as possible

  → “minimum consensus”
- Data management
  - Highly abstracted, following the single-observation concept (van Hintum et al. 1992)
  - Omitting fine-grained metadata

Current EURISCO approach

- EURISCO is increasingly accepted as repository for phenotypic data
- Hundreds of experiments and traits
- But: Data need to be made comparable



Trait Name	Trait Method
Flowering time - beginning	Rating score (1=very early > - 4 (days), 3=early -2 to -4 (days), 5=intermediate +-1 (day), 7=late +2 to +4(days), 9=very late > + 4 (days))
Flowering time	(1,2,3,4,5=5,4,3,2,1 weeks before Bruinsma Wonder(=6) 7,8,9=1,3,5 weeks after)
Flowering time begin	Days after sowing when 50% of plants have opened the first flower(s)
Flowering time begin	(3=early, 7=late)
Flowering time	Gibberellin. Count days from planting to corolla 1st flower visible (1=<41. 2=41-60. 3=61-80. ... 8=161-180. 9=>180)
Flowering time	No treatment. Count days from planting to corolla 1st flower visible (1=<41. 2=41-60. 3=61-80. ... 8=161-180. 9=>180)
Flowering time	count days after 1 May when 50% of florets have opened on 3 flowers
Flowering time	number of days after sowing until first flower head
Flowering time	Count days after 1 April when >50% plants show inflorescence emergence, 999=not flowering during experiment
Flowering time	not vernalized plants: days between sowing and first open flower

1 - 10 of 60

0.14 s

EVA INTRANET

Support of data management

- Provide an intranet platform for project partners
  - Use existing infrastructure for project-specific phenotypic data (in a separate intranet)
    - Exchange format
    - Upload and check tools
  - Provide features for searching/filtering/downloading data
    - Based on users' requirements
  - Extension for privileged access (data embargo period)
    - Data could be published automatically after expiration
    - Automatic requests for approval by NFPs can be implemented
  - Also non-EURISCO material could be managed
    - Handling this data after embargo period needs to be discussed
- Ensure a supportive documentation unit (providing templates, standards, facilitating data flow)

Support of data harmonisation

- Data harmonisation
  - Harmonisation of experiment set-up, treatment etc.
    - Start with minimum approach
      - E.g. MIAPPE (Krajewski et al. 2015)
      - Better description
    - Desirable: harmonised protocols
- Better structuring of traits/methods/scales
  - Support for EVA project partners
    - Support for the development of common vocabularies/approaches
    - Improve comparability
      - Mapping onto ontology terms, e.g. Crop Ontology (Arnaud et al. 2012)
    - Support the mapping process by tools, e.g. suggestion of ontology terms

    → Agreements on common approaches should be the first choice

- Provide training + helpdesk

M. Grau / IPK

THANK YOU FOR YOUR ATTENTION