EURISCO as central information system for data sharing

– Current status and options for EVA –

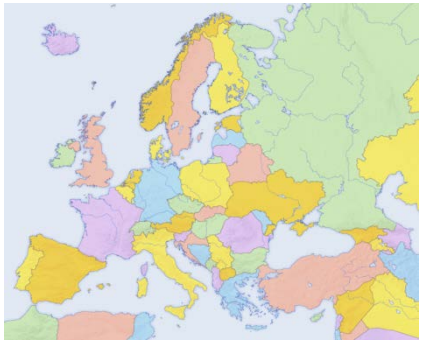
EVA Workshop, 2–3 April 2019, Durrës, Albania

Stephan Weise

Background

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

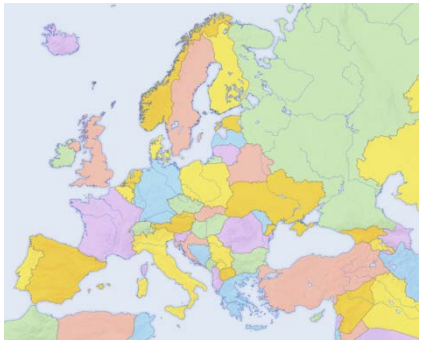
- Purpose
  - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe
  - Assists in meeting national obligations
    - Food and Agriculture Organization of the United Nations (FAO)
    - Convention on Biological Diversity (CBD)
    - International Treaty on Plant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



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

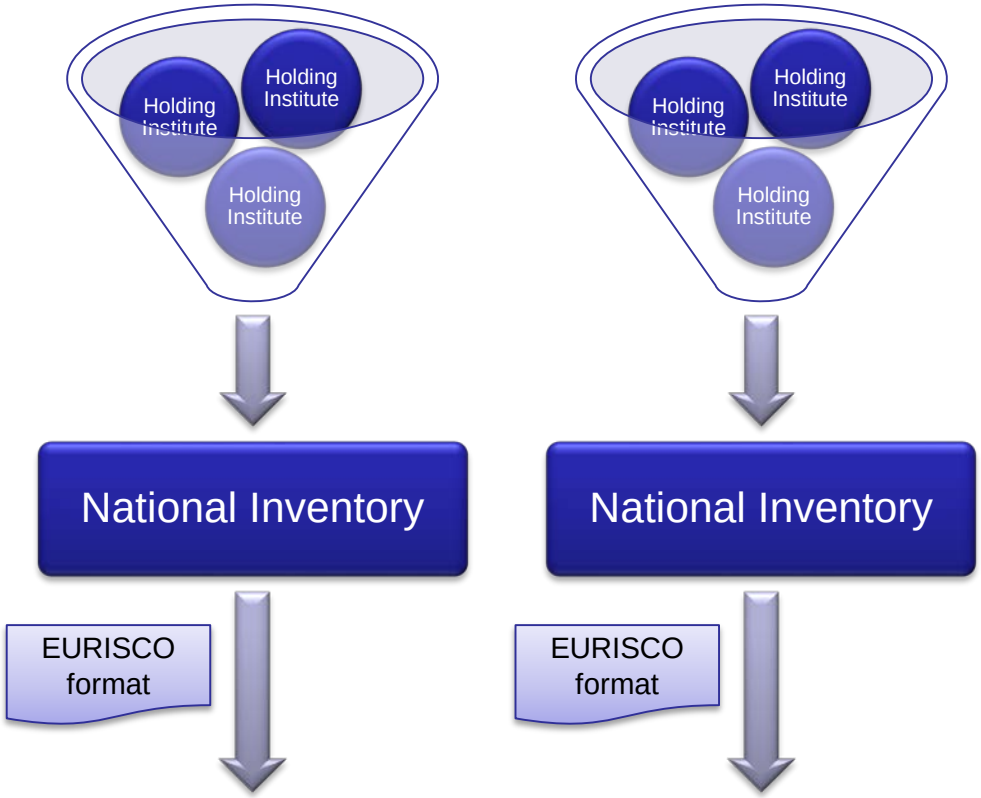
Development

- 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



...

...

383 germplasm collections

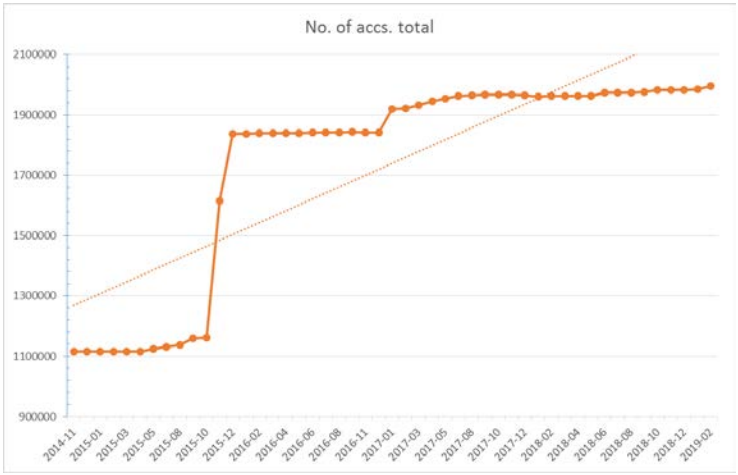
43 member countries

EURISCO

ex situ PGR data

Contents of EURISCO

- 1,995,218 accessions
- 6,405 genera
  (including synonyms, spelling variants)
- 43,495 species names
  (unique combinations genus + species, including synonyms)
- 436,901 MLS accessions
- 56,727 AEGIS accessions
- 39,811 DOIs



as of 2019-03-01

Taxonomic composition

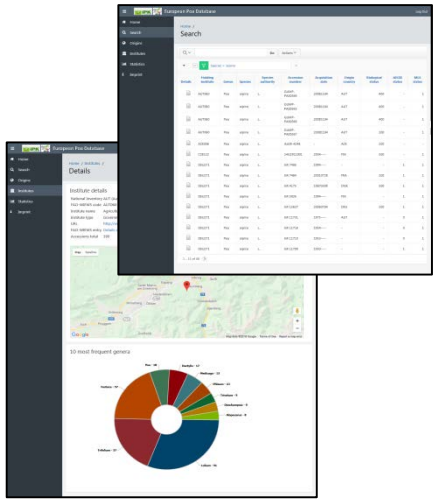
Genus	Species	No. accs.	Total
Arabidopsis	*thaliana*	684,964	685,188
	others	224	
Triticum	*aestivum*	138,516	194,562
(wheat)	*durum*	16,625	
	turgidum	14,816	
	monococcum	1,980	
	spelta	3,232	
	others	19,393	
Hordeum	*vulgare*	107,268	122,512
(barley)	*spontaneum*	5,934	
	others	9,310	
Zea	*mays*	64,988	65,122
(maize)	others	134	
Phaseolus	*vulgaris*	47,069	52,783
(garden bean)	*coccineus*	3,111	
	others	2,603	

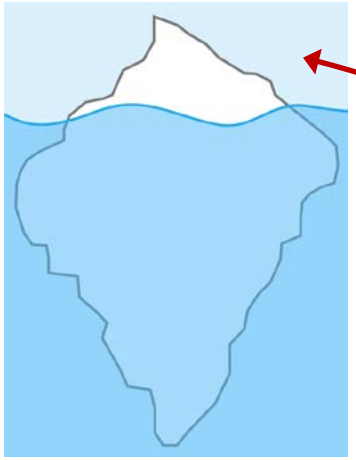
Genus	Species	No. accs.	Total
Solanum	*lycopersicum*	19,977	52,318
(tomato, potato,	*tuberosum*	14,820	
eggplant, etc.)	*andigenum*	2,814	
	melongena	2,131	
	others	12,576	
Avena	*sativa*	33,424	41,561
(oat)	*sterilis*	2,193	
	byzantina	1,976	
	others	3,968	
Pisum	*sativum*	33,275	36,309
(pea)	others	3,034	
Malus	*domestica*	25,166	32,763
(apple)	others	7,597	
Vitis	*vinifera*	25,844	30,197
(grape)	others	4,353	
others			681,903
		Total	1,995,218

as of 2019-03-01

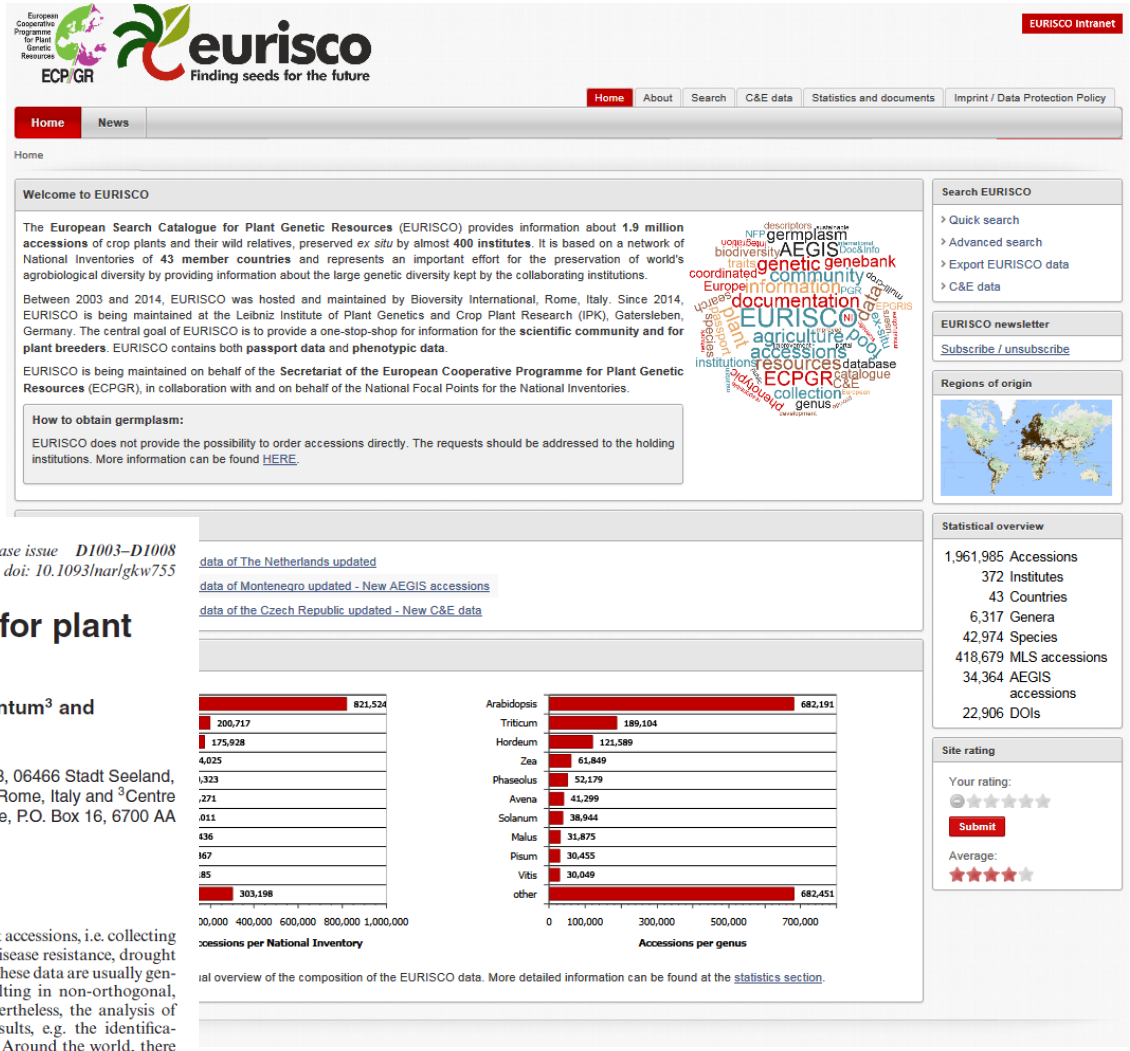
Collaborations

- Genesys (Crop Trust)
  - Regular synchronisation of passport data
  - European hub for Genesys
- FAO-WIEWS
  - Synchronisation of passport data (non-regular)
- Germinate (JHI)
  - Close interlinking
- GLIS (ITPGRFA)
  - EURISCO will provide a service for registering accessions (under implementation)
- ECPGR crop working groups
  - Backend support for crop portals
- AEGIS
  - Flagging via EURISCO





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

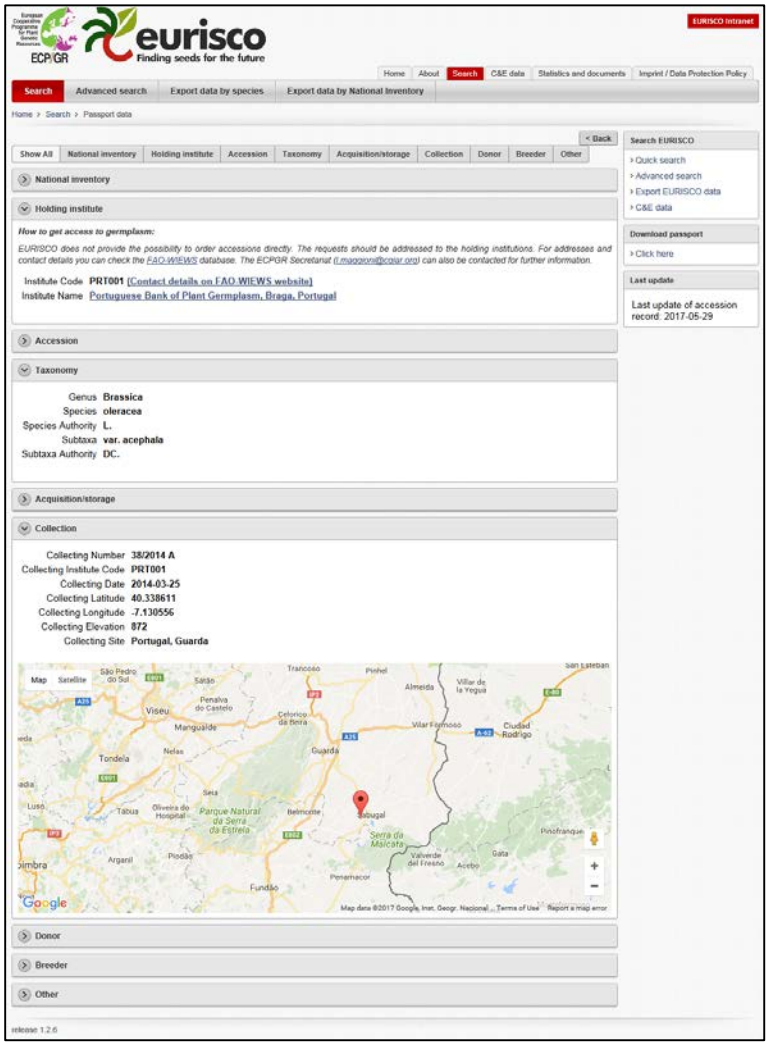
The European Search Catalogue for Plant Genetic Resources, EURISCO, provides information about 1.8 million crop plant accessions preserved by almost 400 institutes in Europe and beyond. EURISCO is being maintained on behalf of the European Cooperative Programme for Plant Genetic Resources. It is based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

46 (sub)versions since 2014

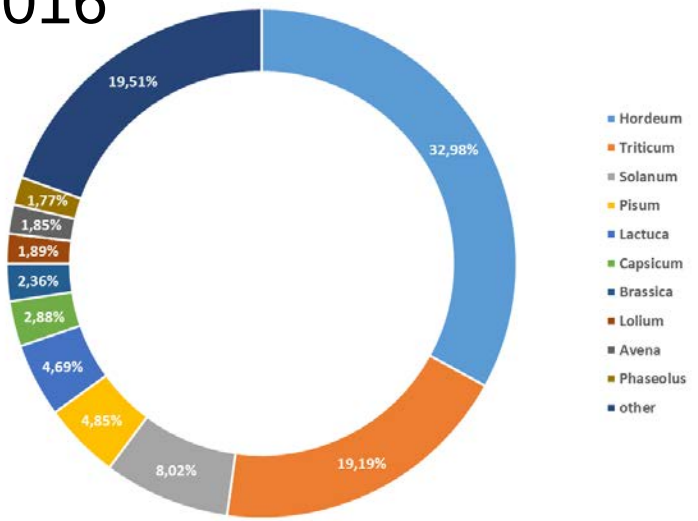
Passport data in EURISCO

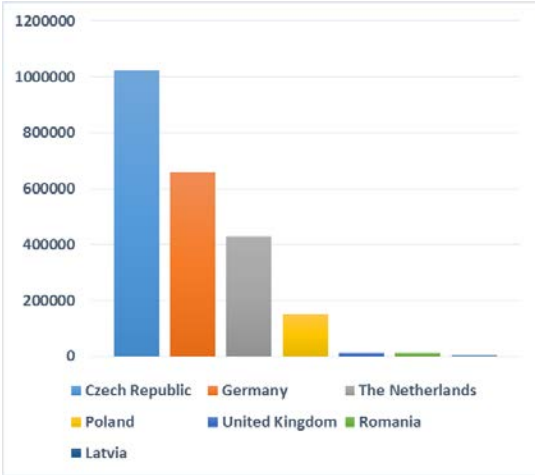
- Four standard searches:
  - Taxonomy (incl. synonyms)
  - Accession
  - Biological status
  - Collecting site
- Advanced search
- Different user-specific export features



Phenotypic data in EURISCO

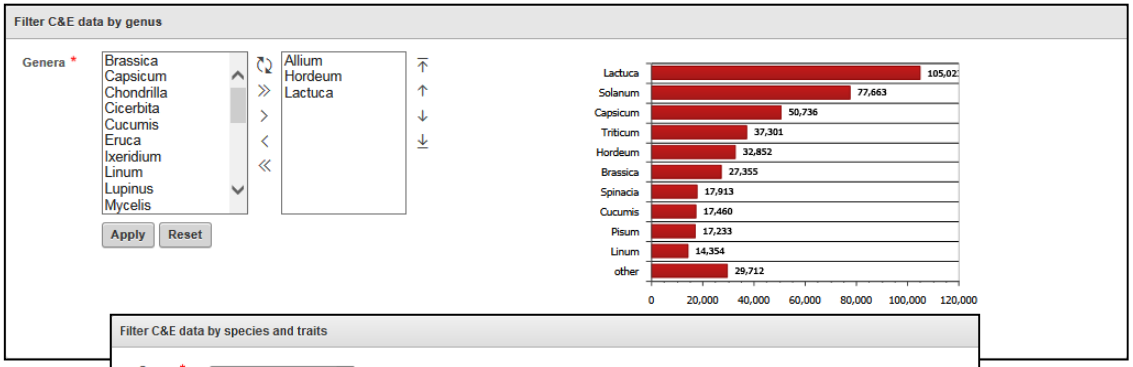
- Extension available since summer 2016
- Currently, 2,286,154 records of data from seven countries
  - Czech Republic
  - Germany
  - Latvia
  - The Netherlands
  - Poland
  - Romania
  - United Kingdom
- 81,759 accs. with phenotypic data

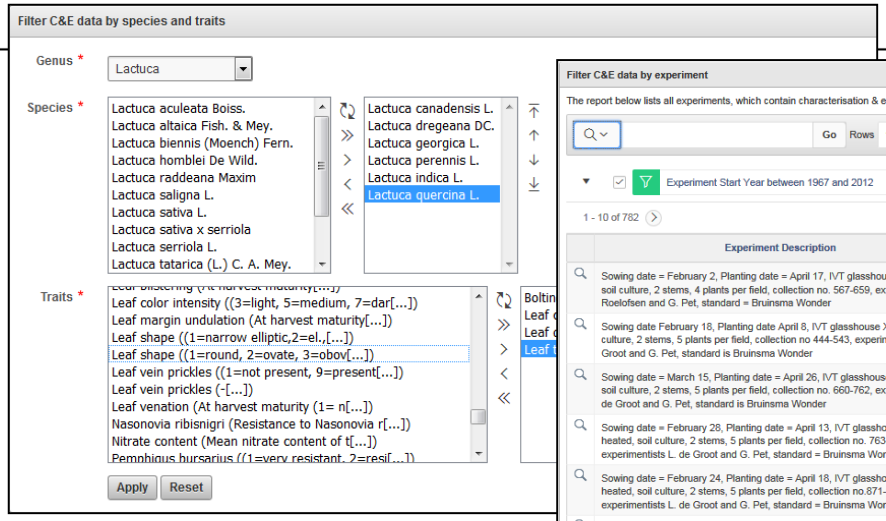


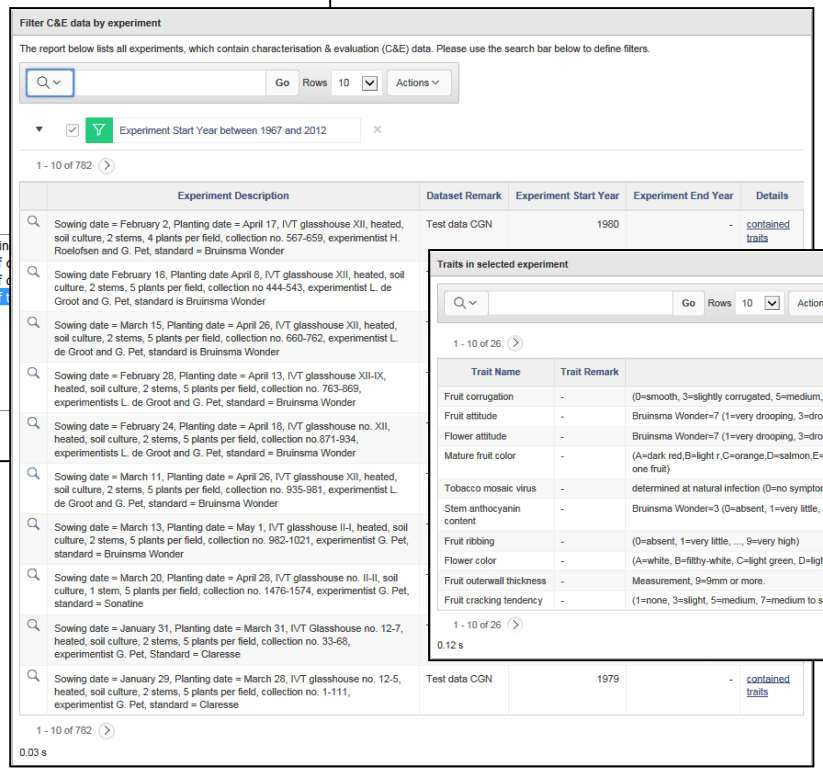


as of 2019-03-29

Search for phenotypic data





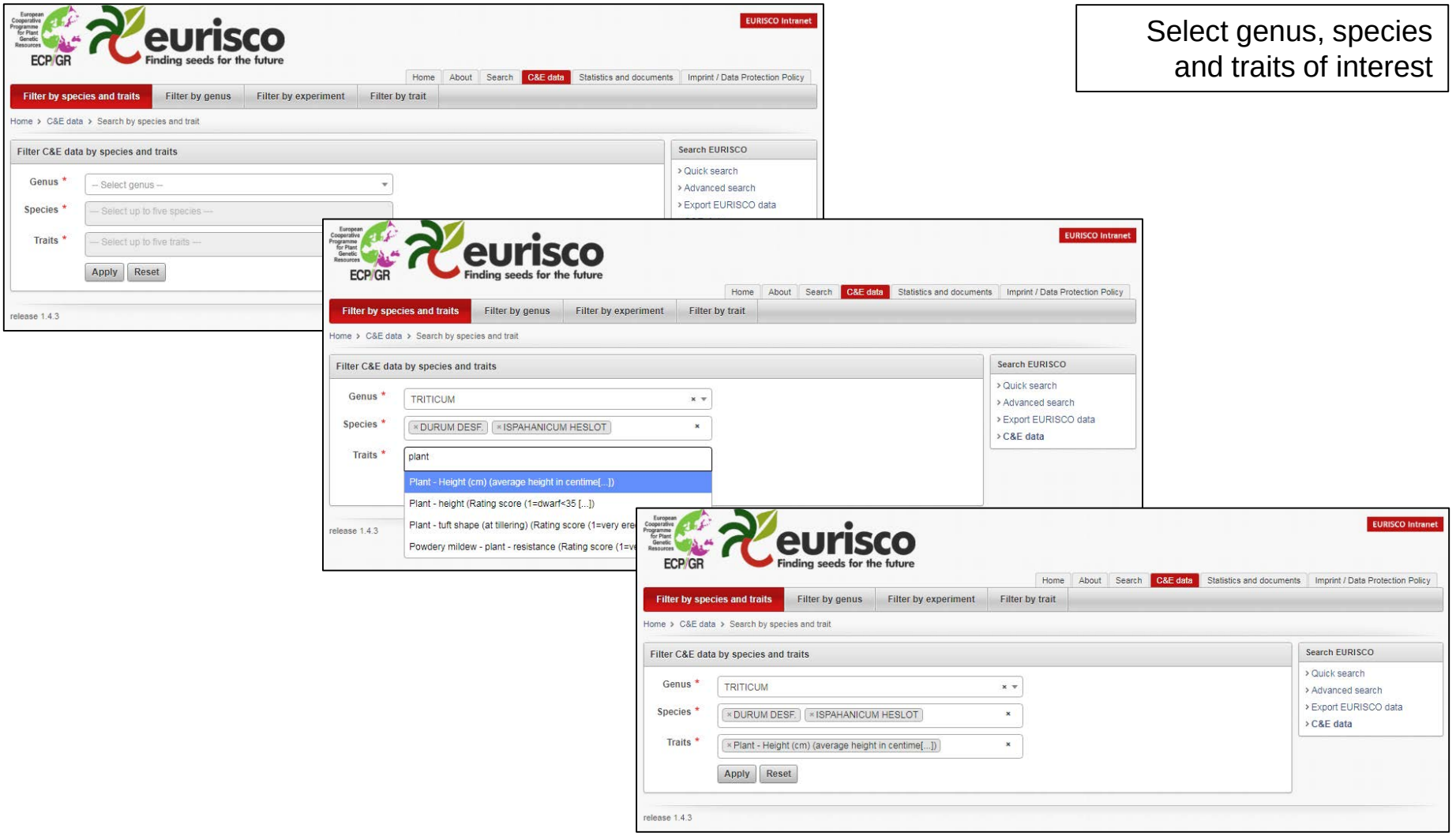


Wizard-based searches for

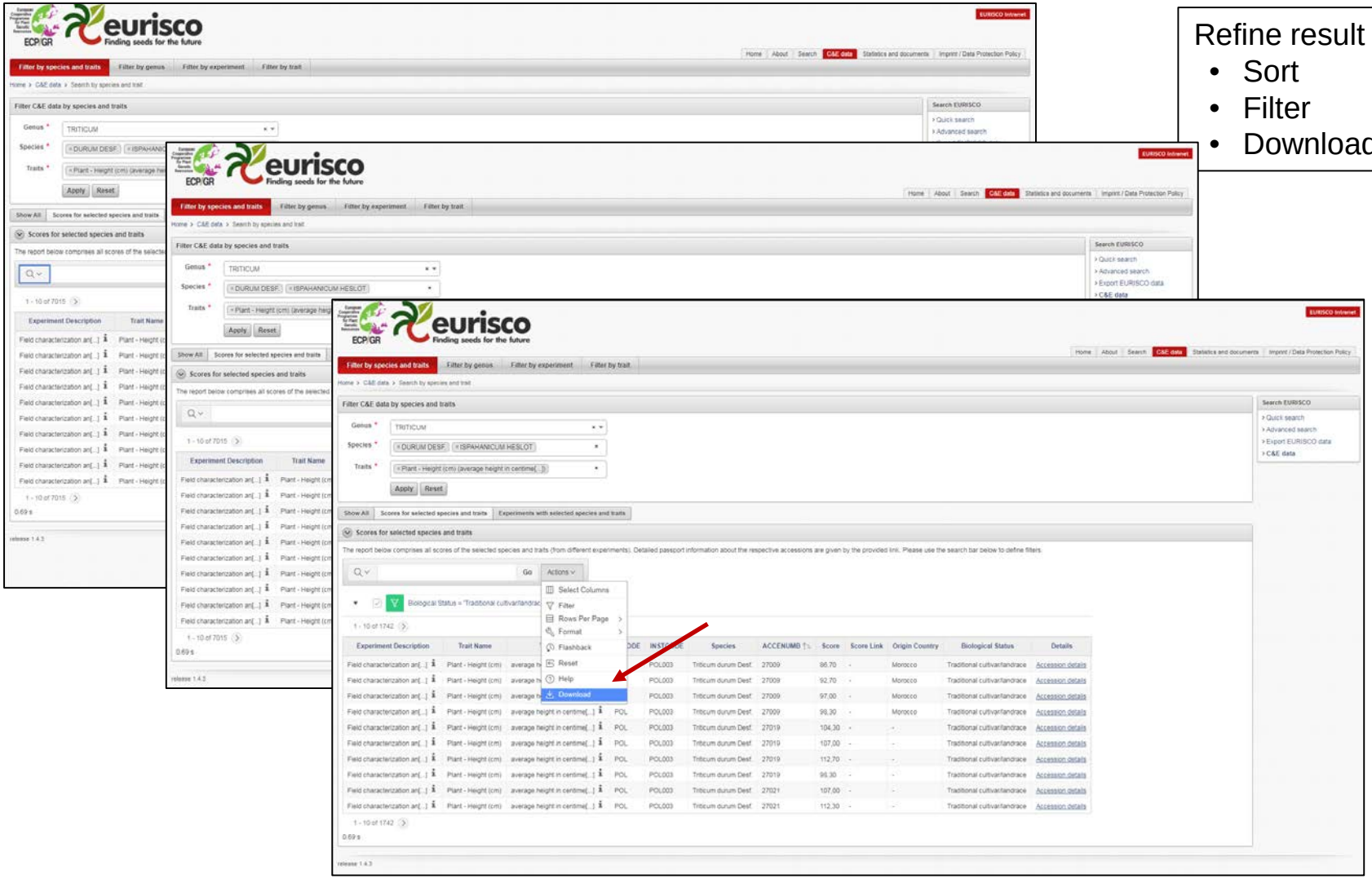
- Genus
- Species and trait
- Experiment
- Trait

Example I – material selection

Select genus, species and traits of interest



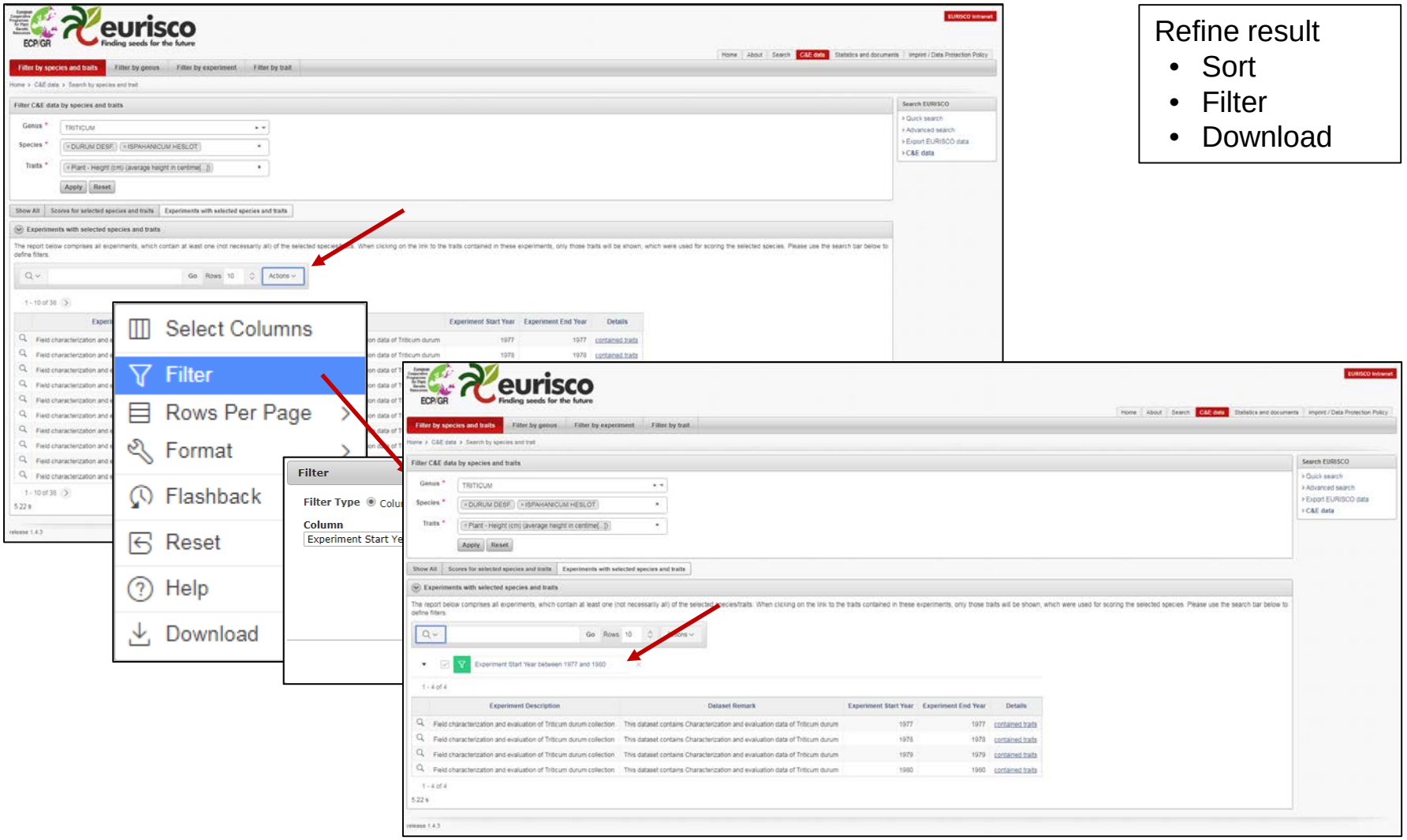
Example I – report of values



Refine result

- Sort
- Filter
- Download

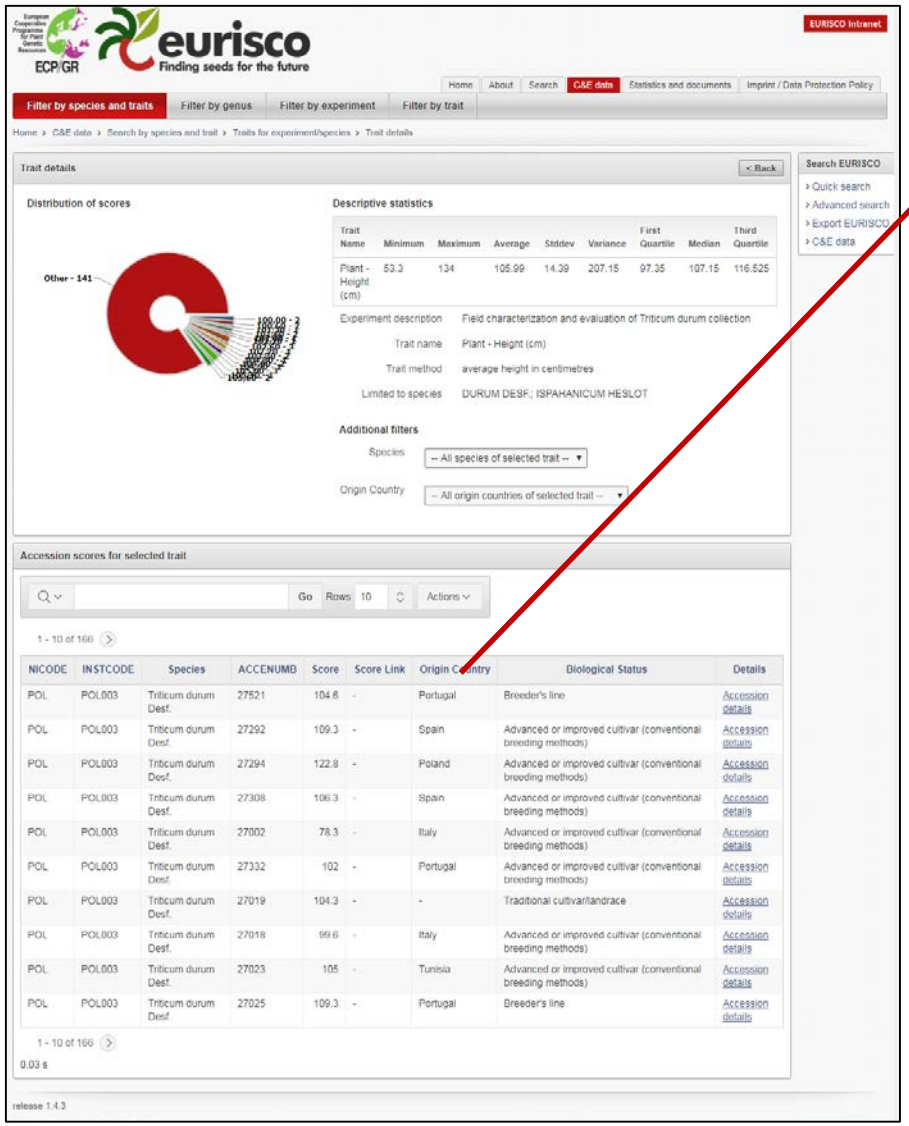
Example I – report of experiments



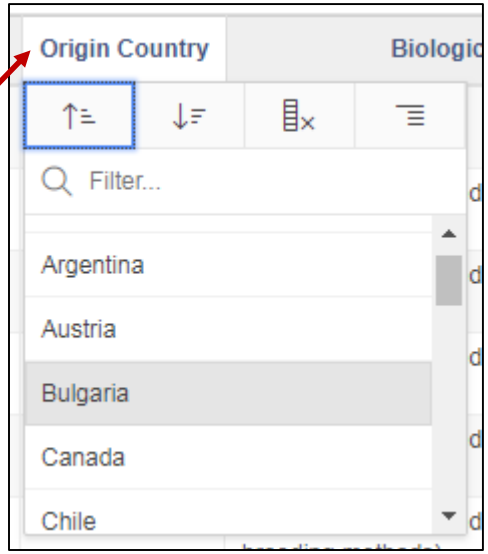
Refine result

- Sort
- Filter
- Download

Example I – scores

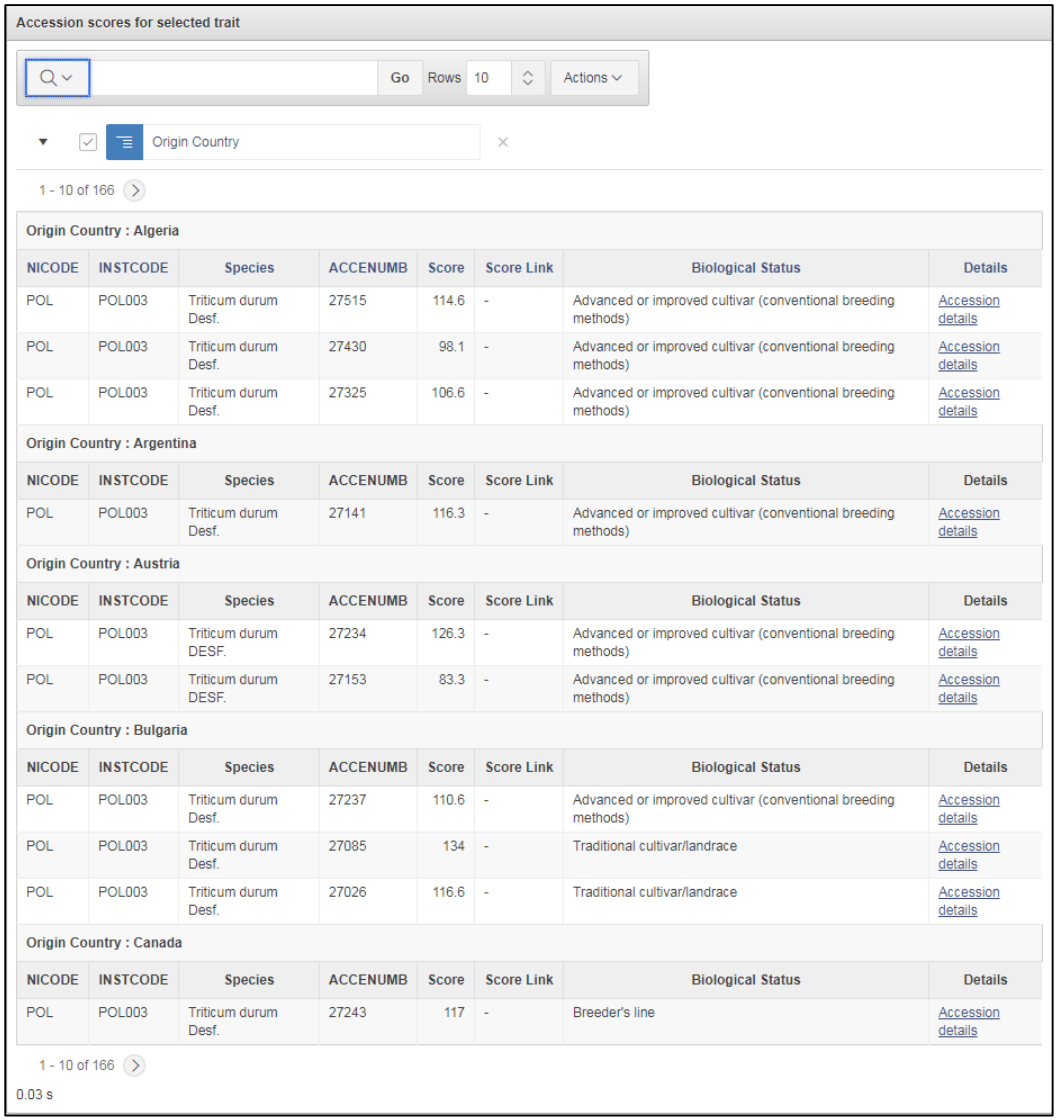


Descriptive statistics



Additional filters, e.g. origin country

Example I – scores



Accession scores for selected trait

Go Rows 10 Actions

Origin Country

1 - 10 of 166

Origin Country : Algeria

NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum Desf.	27515	114.6	-	Advanced or improved cultivar (conventional breeding methods)	Accession details
POL	POL003	Triticum durum Desf.	27430	98.1	-	Advanced or improved cultivar (conventional breeding methods)	Accession details
POL	POL003	Triticum durum Desf.	27325	106.6	-	Advanced or improved cultivar (conventional breeding methods)	Accession details

Origin Country : Argentina

NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum Desf.	27141	116.3	-	Advanced or improved cultivar (conventional breeding methods)	Accession details

Origin Country : Austria

NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum DESF.	27234	126.3	-	Advanced or improved cultivar (conventional breeding methods)	Accession details
POL	POL003	Triticum durum DESF.	27153	83.3	-	Advanced or improved cultivar (conventional breeding methods)	Accession details

Origin Country : Bulgaria

NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum Desf.	27237	110.6	-	Advanced or improved cultivar (conventional breeding methods)	Accession details
POL	POL003	Triticum durum Desf.	27085	134	-	Traditional cultivar/landrace	Accession details
POL	POL003	Triticum durum Desf.	27026	116.6	-	Traditional cultivar/landrace	Accession details

Origin Country : Canada

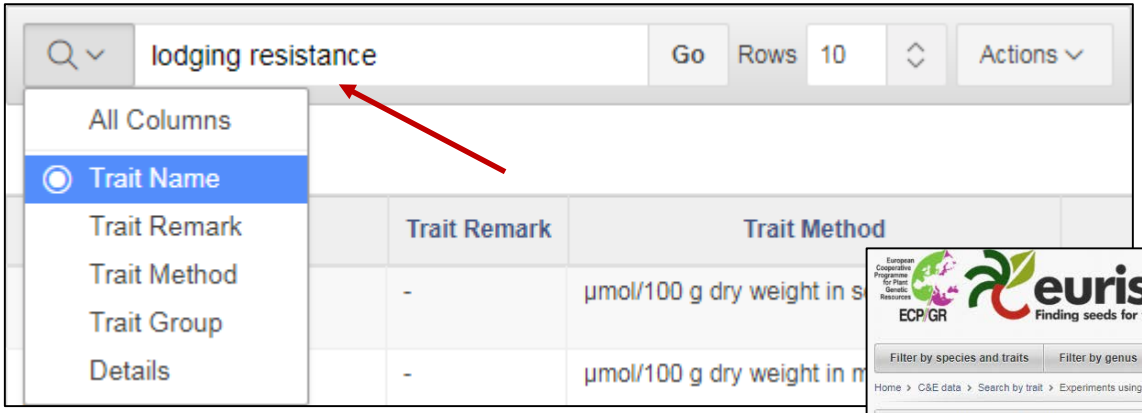
NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum Desf.	27243	117	-	Breeder's line	Accession details

1 - 10 of 166

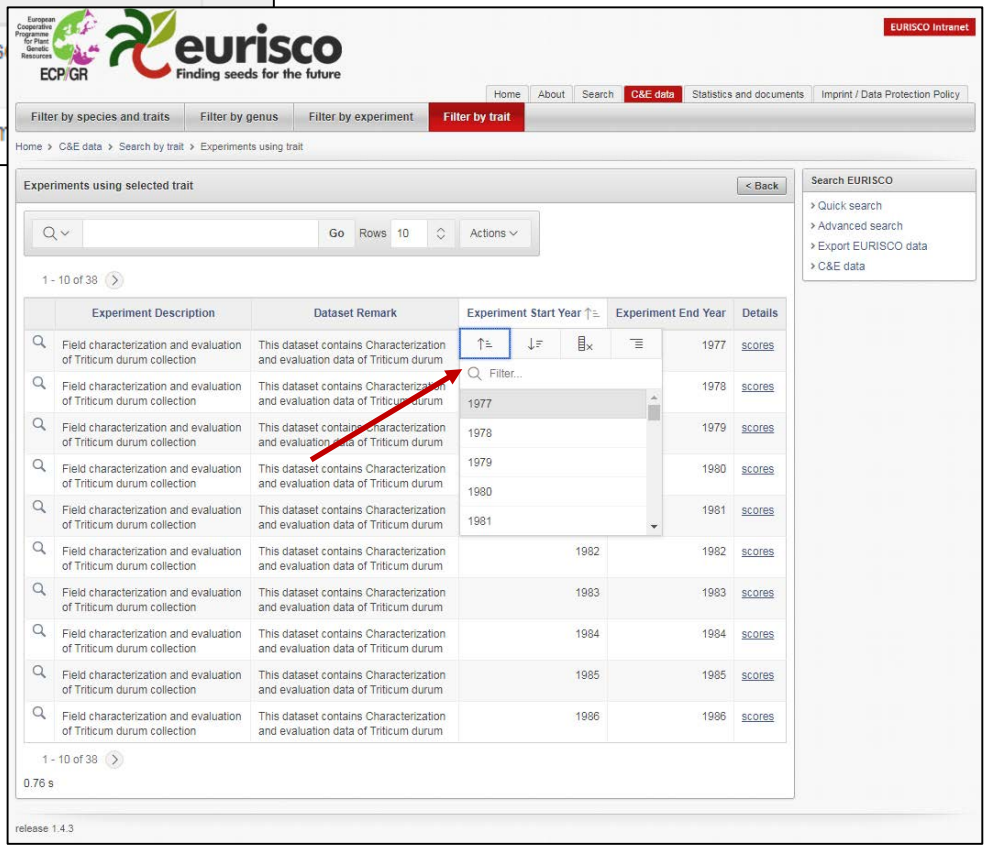
0.03 s

Group values

Example II – trait selection

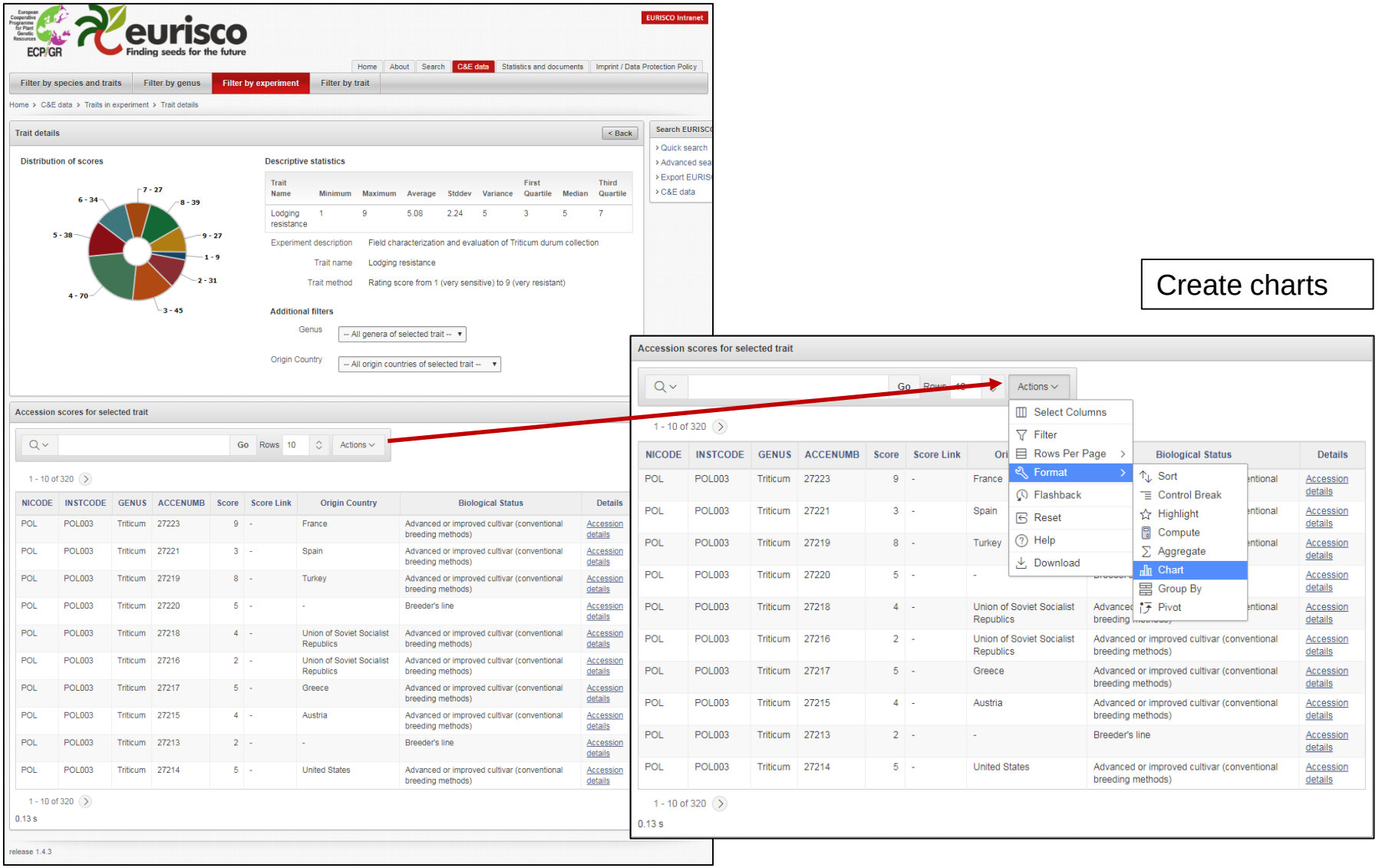


Search for trait of interest



Filter experiments containing the selected trait

Example II – scores



Create charts

Example II – scores

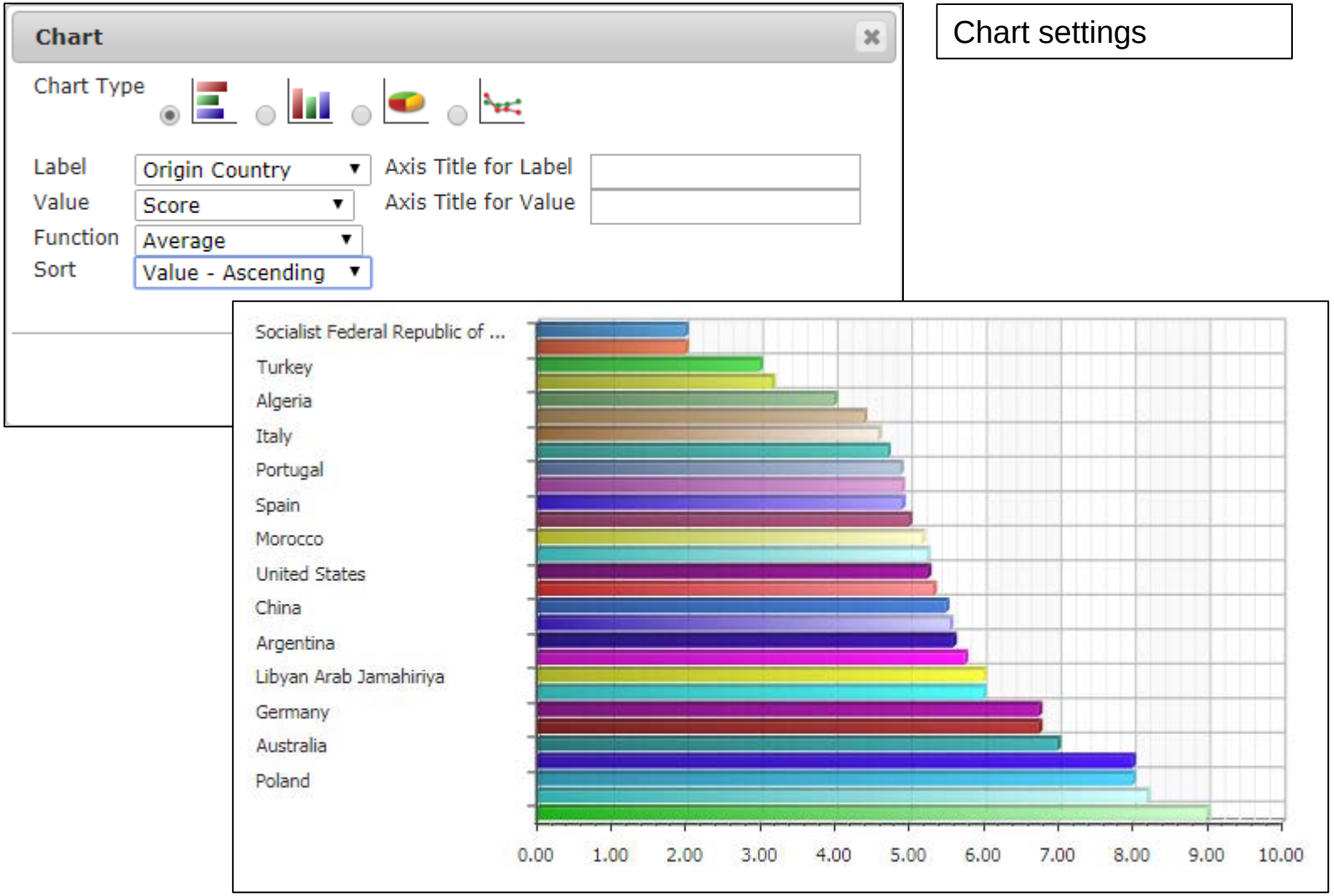
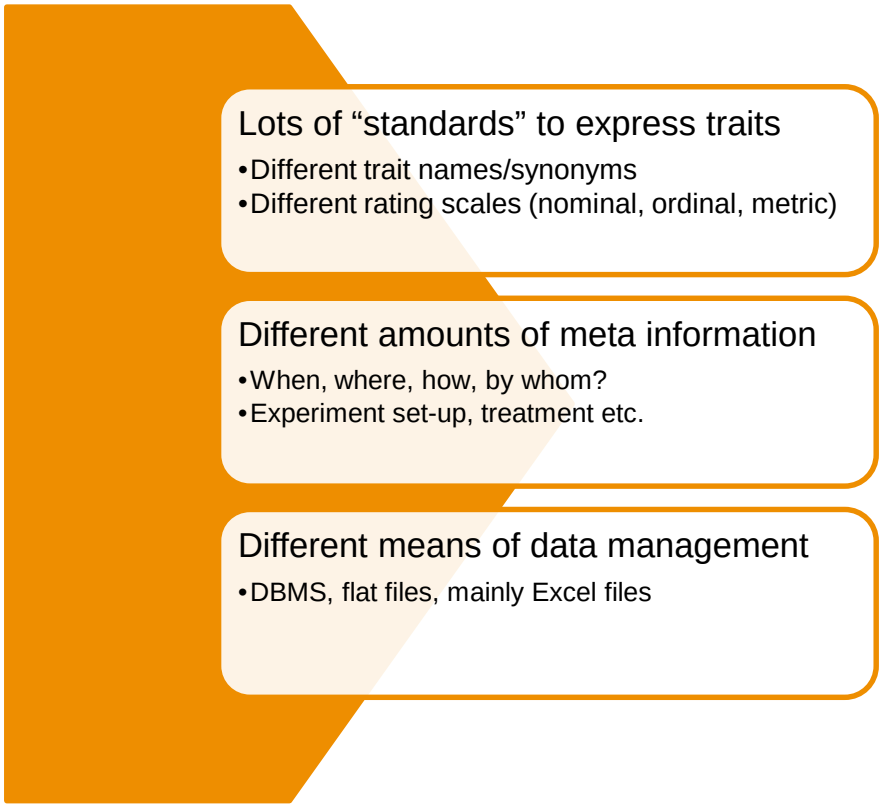


Chart settings

The challenge: Diversity of data



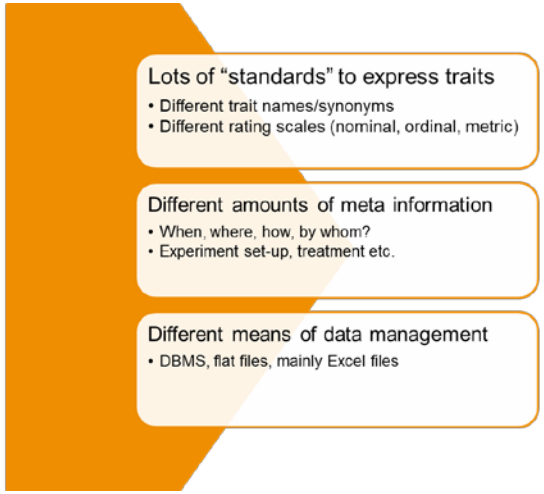
Challenges for EURISCO

- Use of phenotypic traits
  - Lots of different traits, methods, scales
  - About 600 germplasm collections in Europe, around 400 in EURISCO

  → ***Unlikely to get to a standardisation***

- Information technology diversity
  - Different data management systems (DBMS, flat files, Excel)
  - Different level of IT support
  - Varying IT affinity
    → problem of acceptance of sophisticated solutions

→ But needed: one-size-fits-all method



Current 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

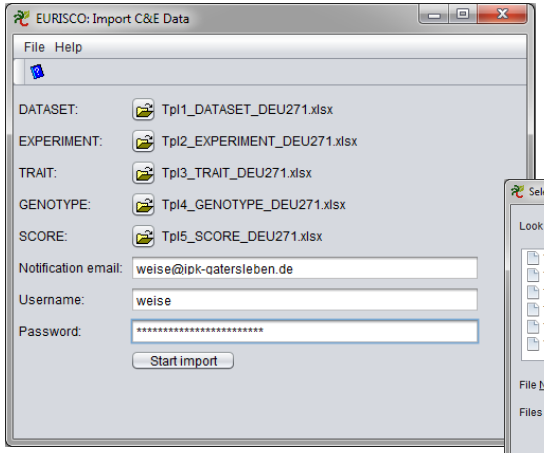
  → “minimum consensus”
- Data management
  - Highly abstracted, following the single-observation concept (van Hintum et al. 1992)
  - Omitting fine-grained metadata

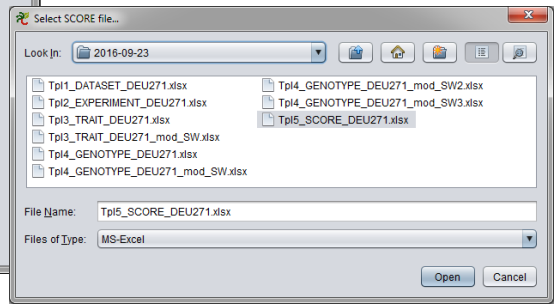
Data upload in three steps

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*





Current limitations

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

What could be improved?

- Through an EVA project, it could be possible to support:
  - 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 crop experts (ECPGR crop working groups)
      - Focus on most active groups at the beginning
      - 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

What could be improved?

- Through an EVA project, it could be possible to:
  - 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
    - 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)

M. Grau / IPK

THANK YOU FOR YOUR ATTENTION