Documentation of European Plant Genetic Resources

The EURISCO information system

First Meeting of the ECPGR Maize Working Group,
2nd to 3rd December 2019, Belgrade, Serbia

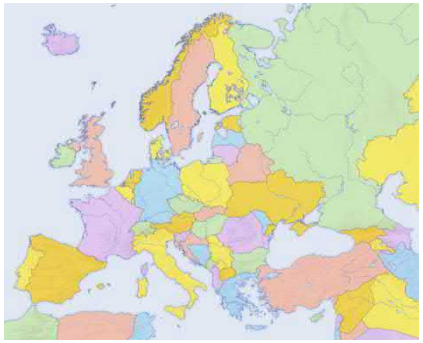
Stephan Weise
2 December 2019

INTRODUCTION AND BACKGROUND

Background

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

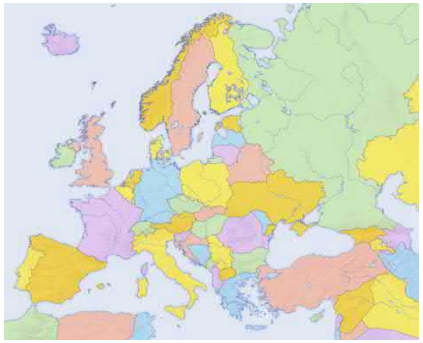
- Purpose
  - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe
  - Assists in meeting national obligations
    - Food and Agriculture Organization of the United Nations (FAO)
    - Convention on Biological Diversity (CBD)
    - International Treaty on Plant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



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

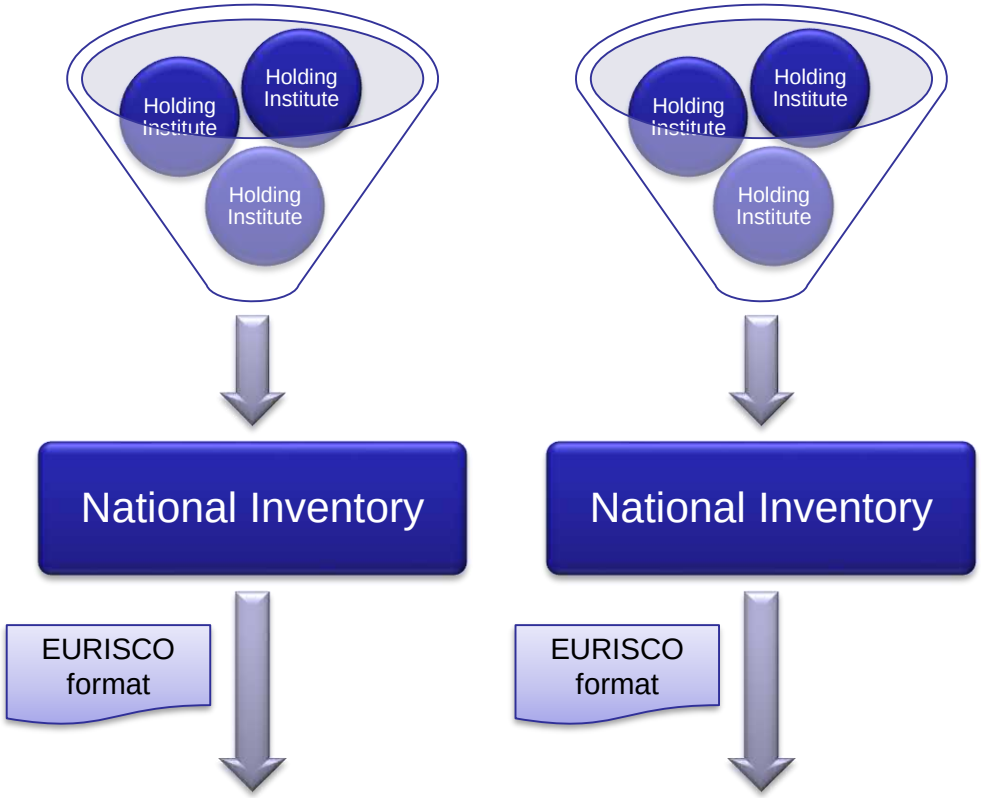
Development

- 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



...

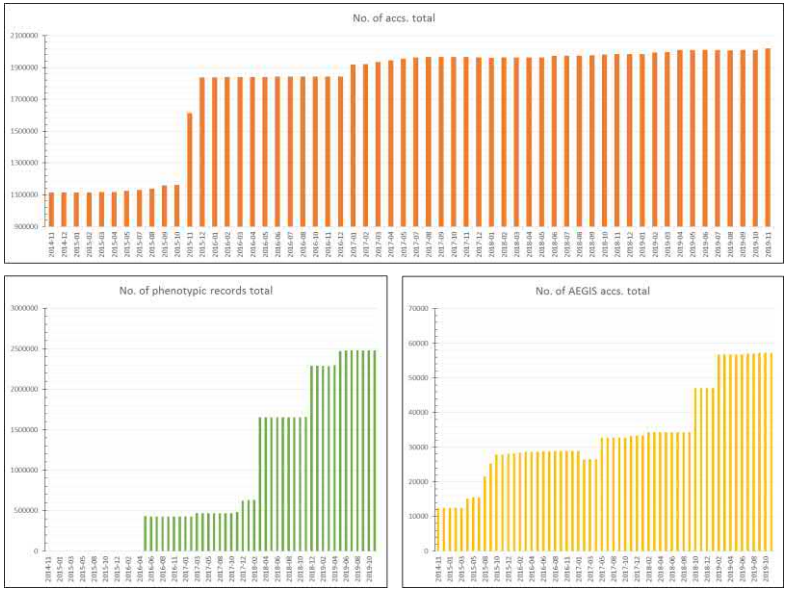
398 germplasm collections

...

43 member countries

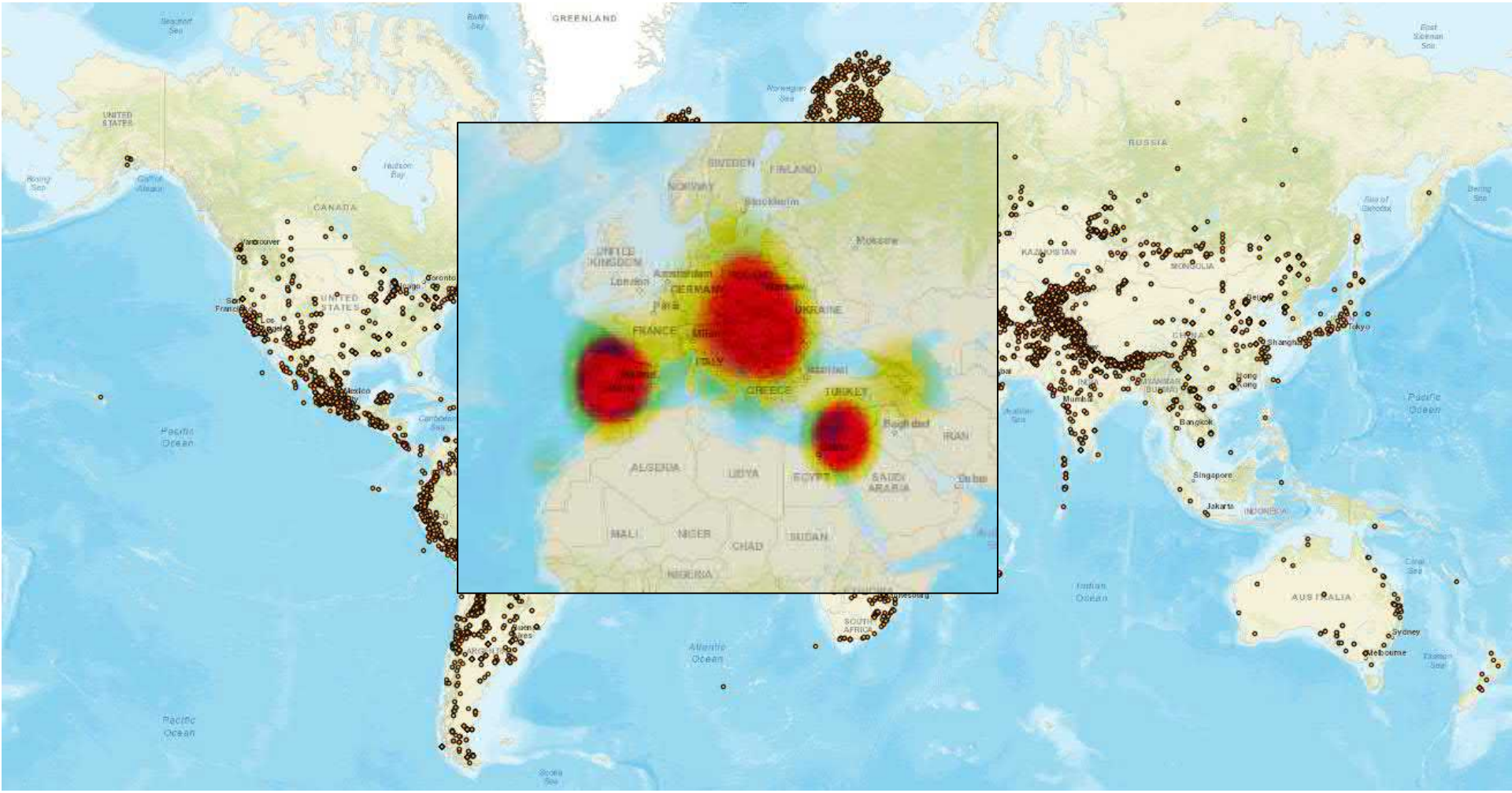
Contents of EURISCO

- 2,019,404 accessions
- 6,392 genera
  (including synonyms, spelling variants)
- 43,223 species names
- 444,086 MLS accessions
- 57,294 AEGIS accessions
- 60,490 DOIs



as of 2019-11-25

Collecting sites



as of 2017-05-31

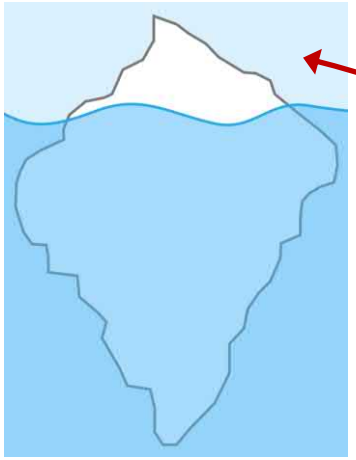
Taxonomic composition

Genus	Species	No. accs.	Total
Arabidopsis	*thaliana*	684,964	685,188
	others	224	
Triticum	*aestivum*	140,133	196,495
(wheat)	*durum*	16,857	
	turgidum	14,860	
	monococcum	1,986	
	spelta	3,260	
	others	19,399	
Hordeum	*vulgare*	113,649	123,121
(barley)	*spontaneum*	6,164	
	others	3,308	
Zea	*mays*	65,560	65,686
(maize)	others	126	
Phaseolus	*vulgaris*	47,283	53,133
(garden bean)	*coccineus*	3,177	
	others	2,673	

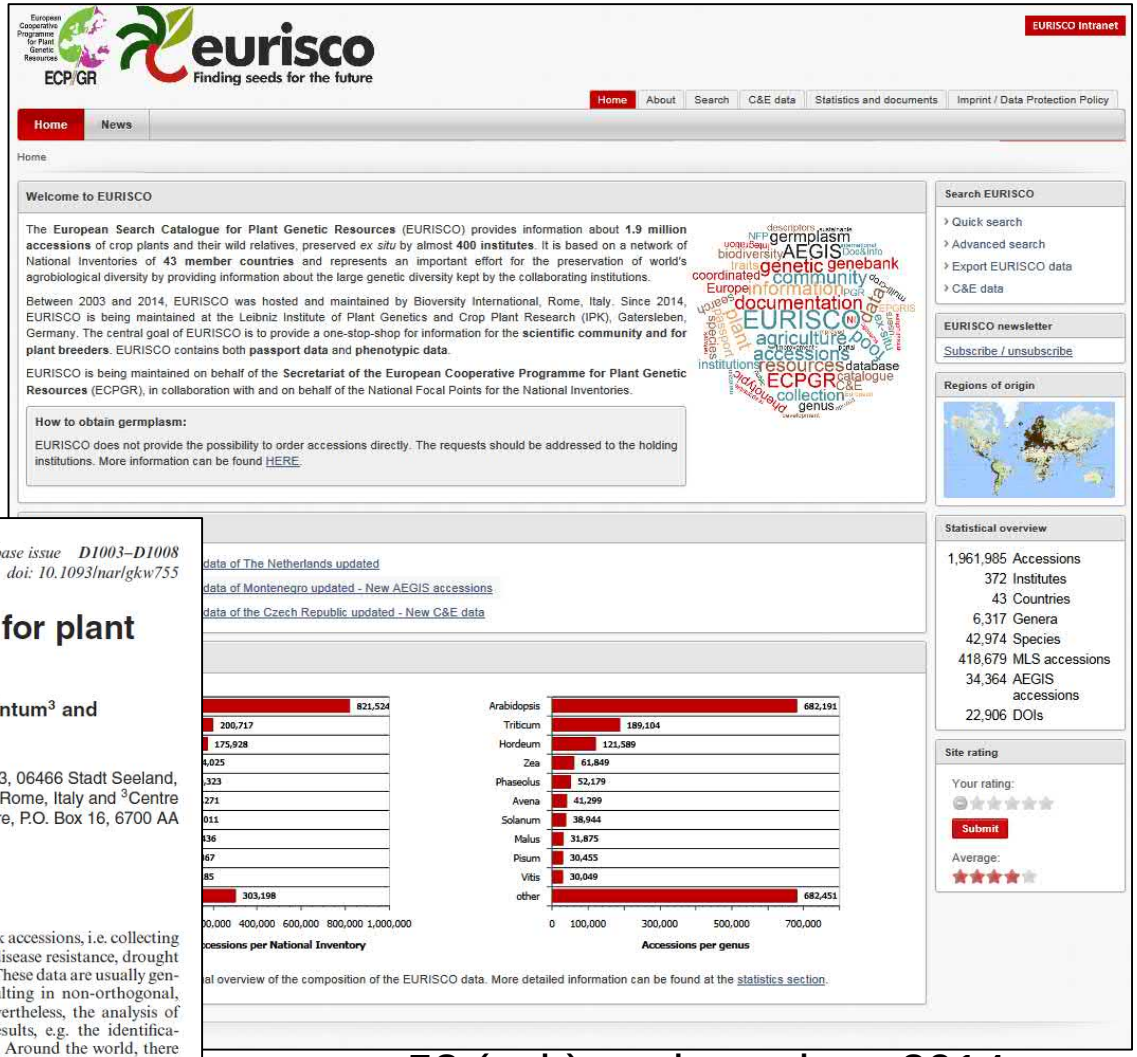
Genus	Species	No. accs.	Total
Solanum	*lycopersicum*	20,468	50,619
(tomato, potato,	*tuberosum*	14,912	
eggplant, etc.)	*andigenum*	2,814	
	melongena	2,125	
	others	10,300	
Vitis	*vinifera*	35,335	41,902
(grape)	others	6,567	
Avena	*sativa*	33,561	41,774
(oat)	*sterilis*	2,203	
	byzantina	1,987	
	others	4,023	
Pisum	*sativum*	33,474	36,539
(pea)	others	3,065	
Malus	*domestica*	31,553	33,504
(apple)	others	1,951	
others			691,443
		Total	2,019,404

as of 2019-11-25

EURISCO WEB



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

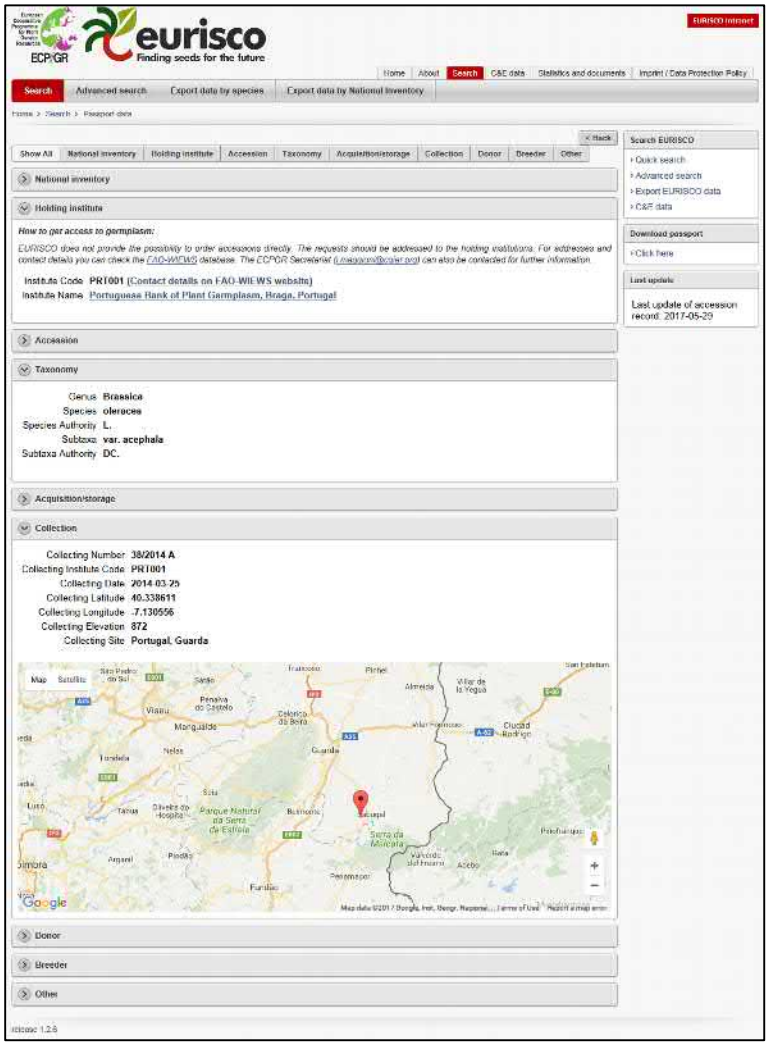
The European Search Catalogue for Plant Genetic Resources, EURISCO, provides information about 1.8 million crop plant accessions preserved by almost 400 institutes in Europe and beyond. EURISCO is being maintained on behalf of the European Cooperative Programme for Plant Genetic Resources. It is based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

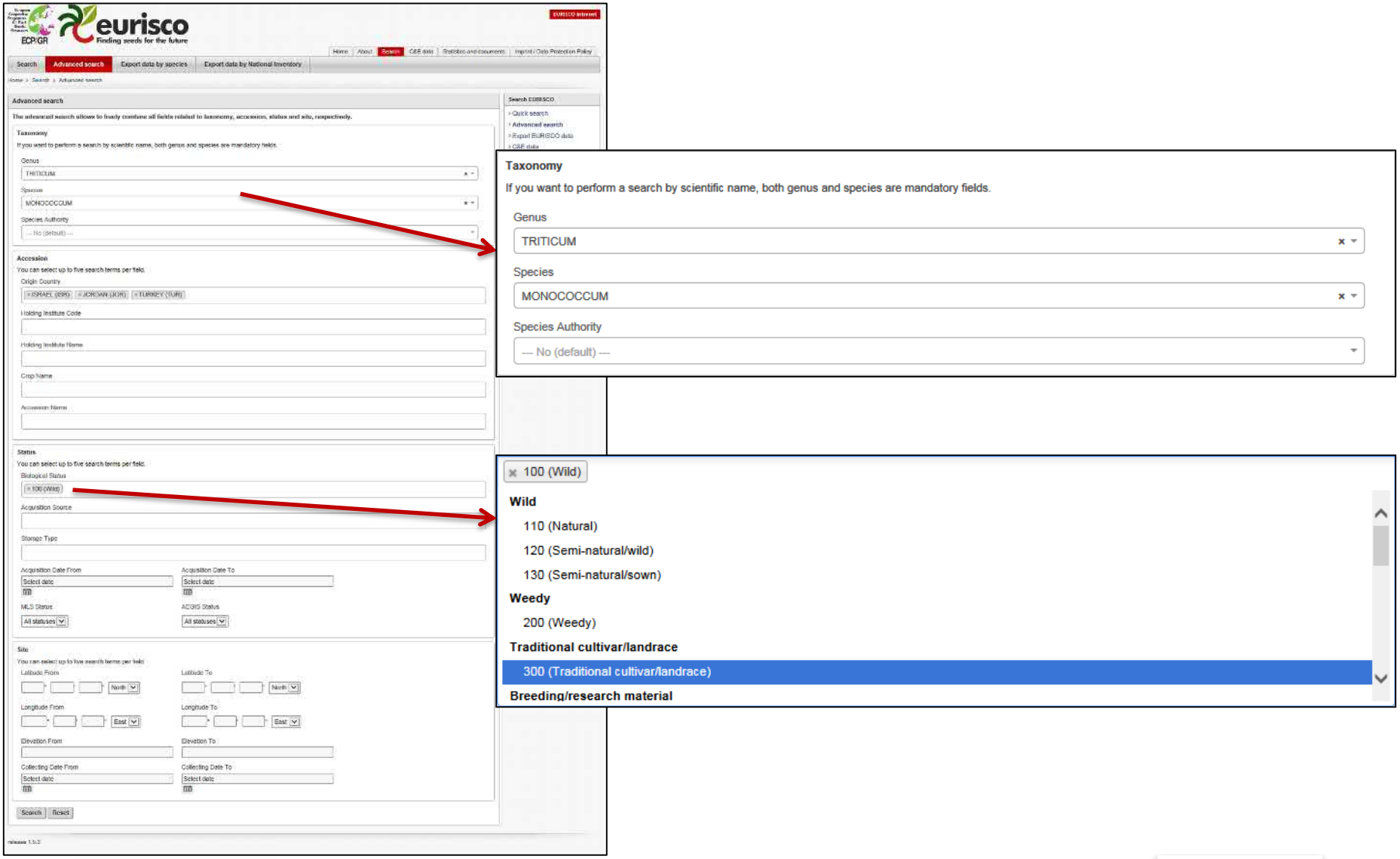
53 (sub)versions since 2014

Passport data search in EURISCO

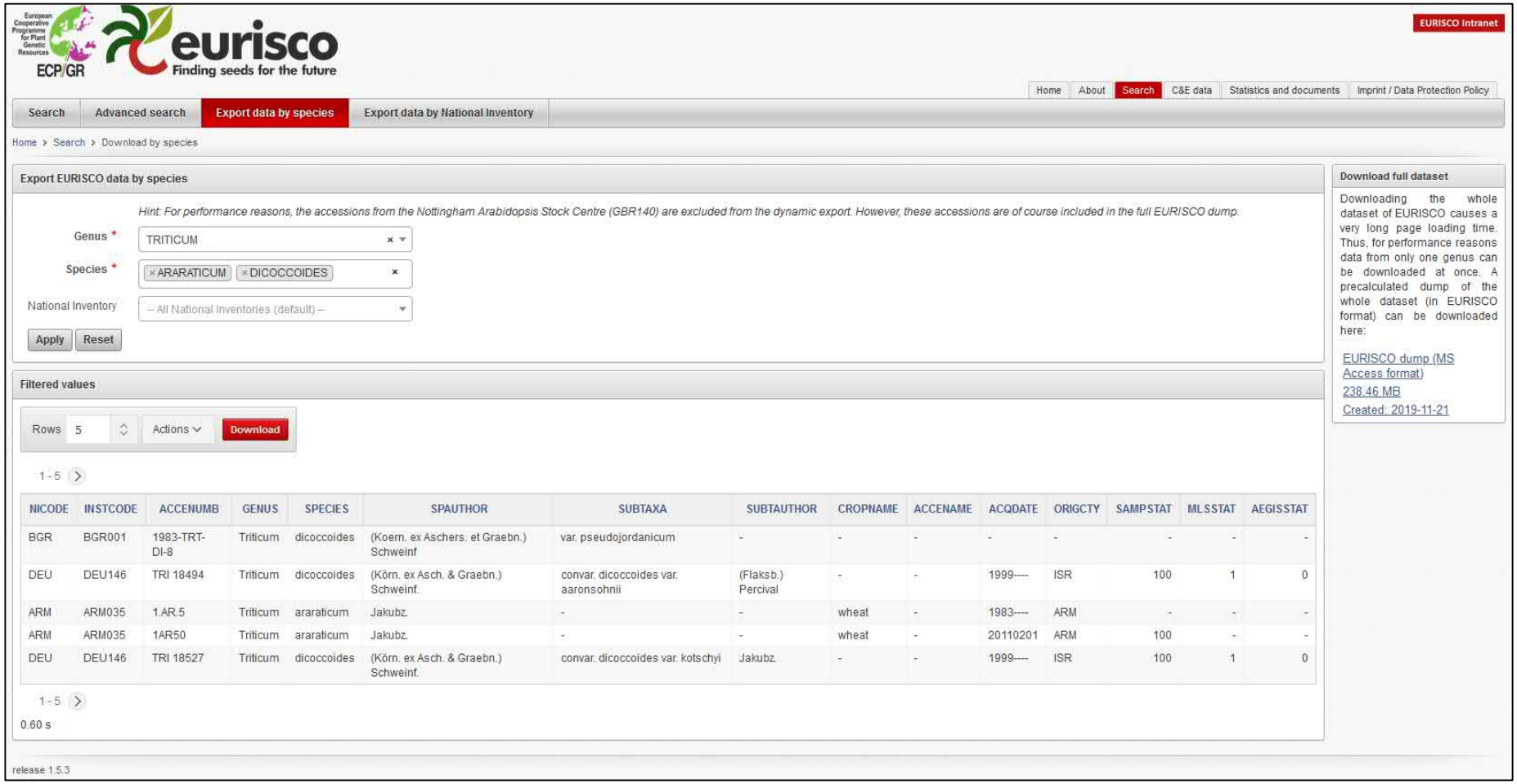
- Four standard searches:
  - Taxonomy (incl. synonyms)
  - Accession
  - Biological status
  - Collecting site
- Advanced search
- Different user-specific export features



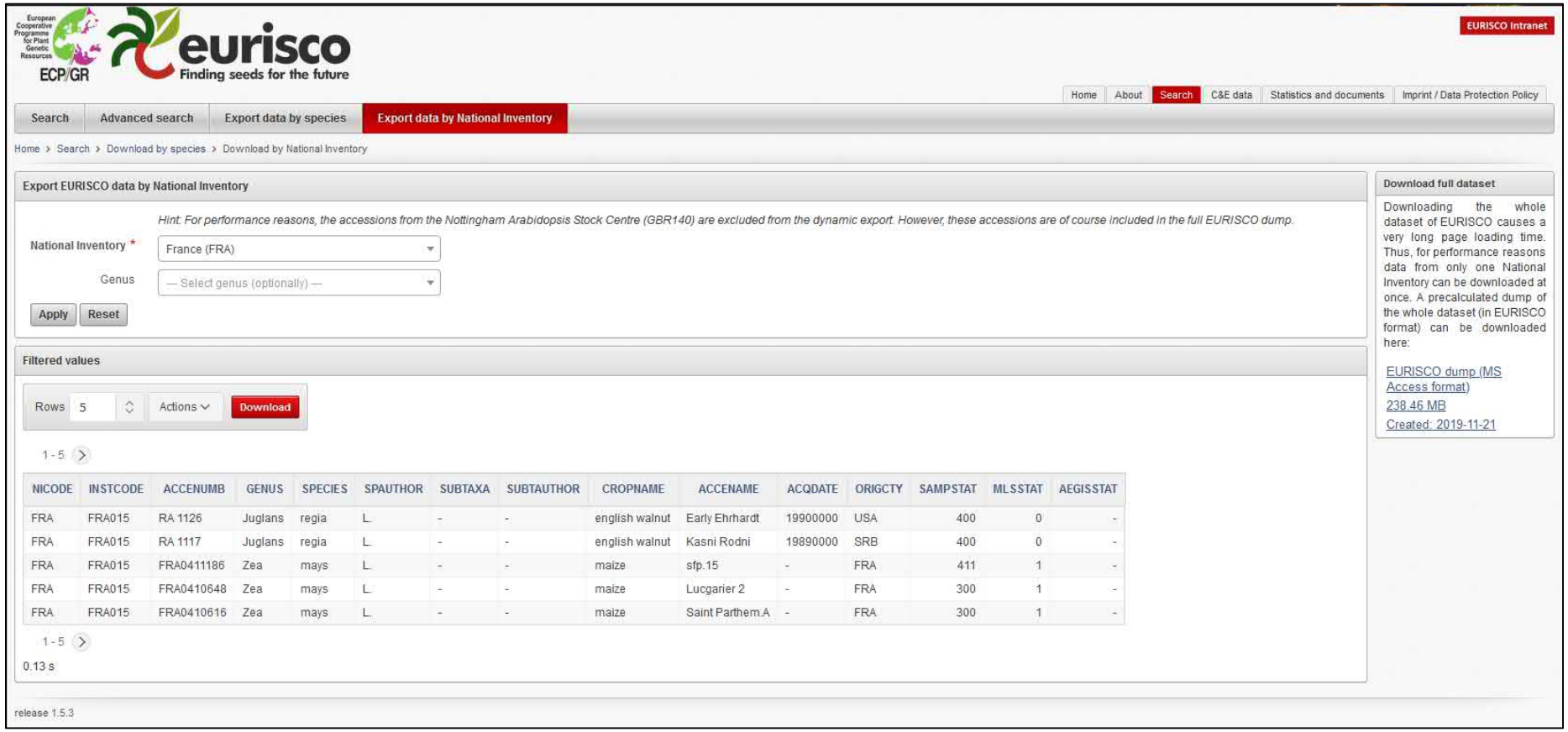
Advanced search form



User specific export – by species



User specific export – by NI

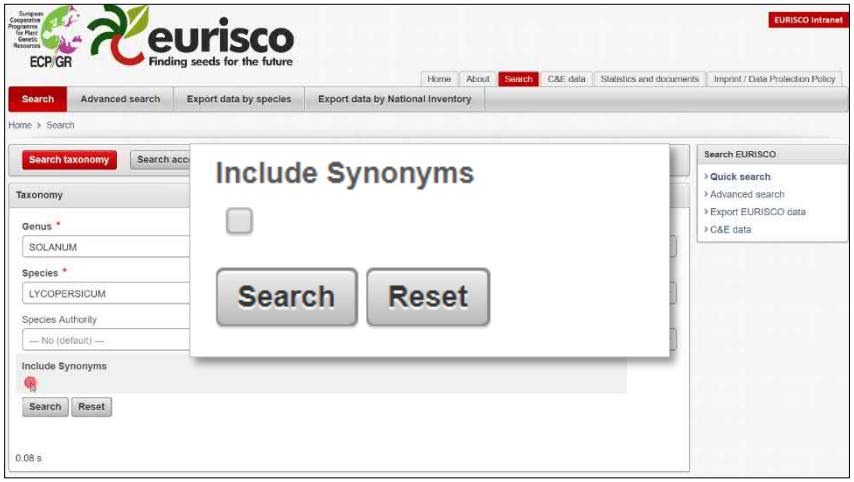


Taxonomy search

- Challenges
  - Data from almost 400 institutes
    - Different taxonomic schools, opinions, traditions
    - No uniform scientific names, e.g. different author abbreviations
    - Misspellings
    - No curation in EURISCO (only by data provider) → DSA
  - Knowledge on data background needed for searches
- Improvement of taxonomic search
  - Identification of synonym candidates
    - GRIN, Mansfeld
      + taxonomic terms accepted in EURISCO (based on user feedback)
  - Obtaining more complete search results
  - However, limited to available data!

Taxonomy search

- Rework of the search interface
  - Allows to include synonyms
  - Shows the distribution of matching types
  - Significant performance improvement



© NIAB 2019
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization; 1–3
doi:10.1017/S1479262119000339

Short Communication

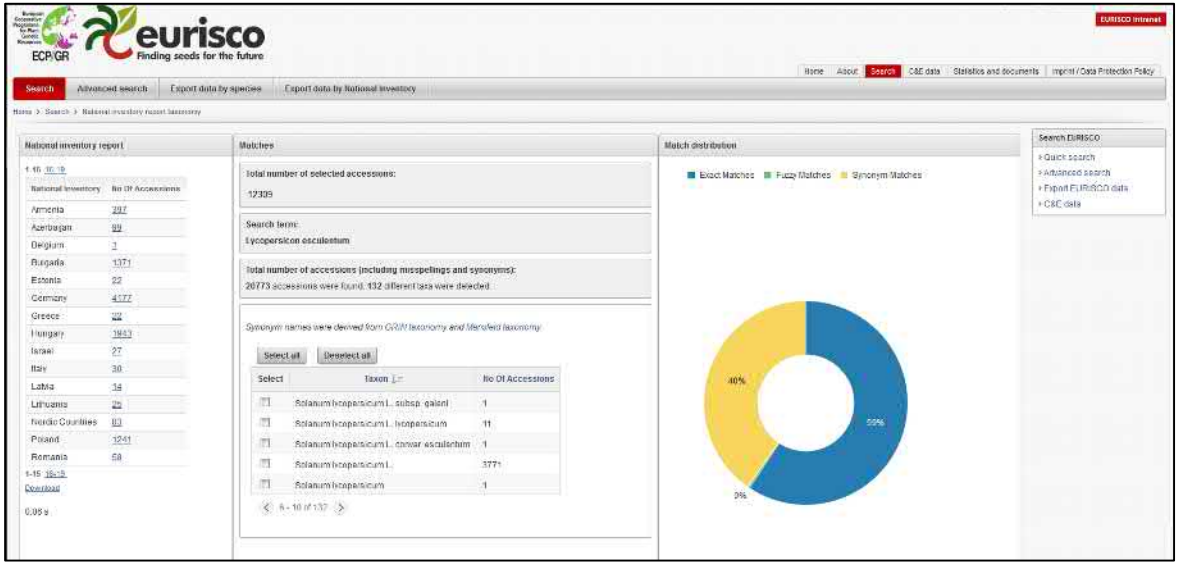
Advancement of taxonomic searches in the European search catalogue for plant genetic resources

Stefanie Kreide* , Markus Oppermann and Stephan Weise
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Seeland, Germany

Received 6 September 2019; Accepted 4 October 2019

Abstract

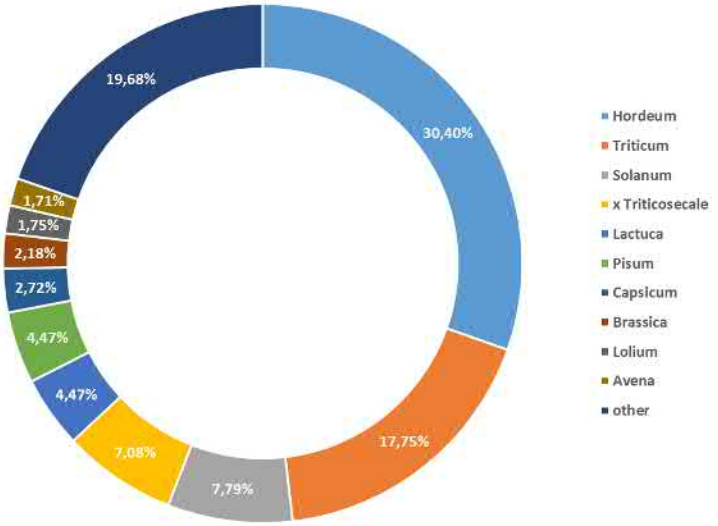
Genebanks play an important role in the conservation of global plant biodiversity. The European Search Catalogue for Plant Genetic Resources (EURISCO) was created as a central entry point to provide information on these collections. However, a major challenge lies in the heterogeneity of scientific plant names. This makes the selection of suitable plant material, e.g. for research or breeding purposes, significantly more difficult. For this reason, the taxonomic backbone of EURISCO has been completely revised. Search terms entered by users are now automatically checked against taxonomic reference repositories, allowing a variety of synonyms to be identified. In addition, a fuzzy search has been implemented, which makes the search function tolerant of erroneous data (e.g. caused by typing errors). Besides improvements of the search interface, more support will be given to EURISCO's data providers. The new developments provide a tool that makes it easier to identify problem cases within the data, such as accepted/non-accepted taxonomic names, and

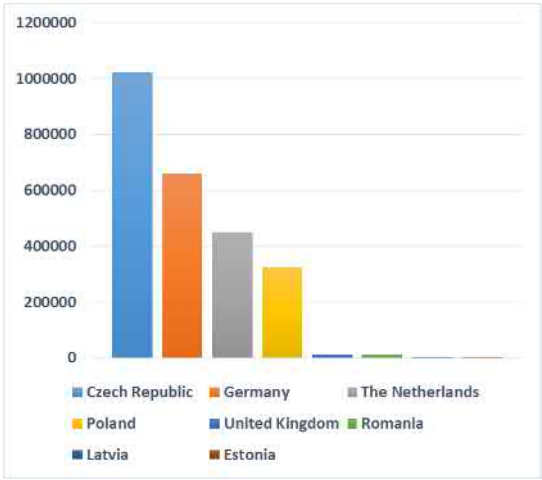


PHENOTYPIC DATA

Phenotypic data

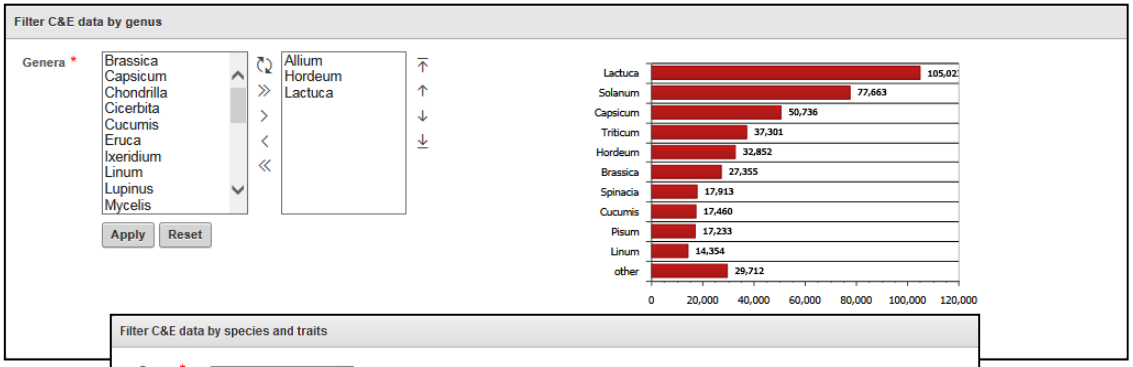
- Extension available since 2016
- Currently, 2,482,274 records of data from eight countries
  - Czech Republic
  - Estonia
  - Germany
  - Latvia
  - The Netherlands
  - Poland
  - Romania
  - United Kingdom
- 84,433 accs. with phenotypic data

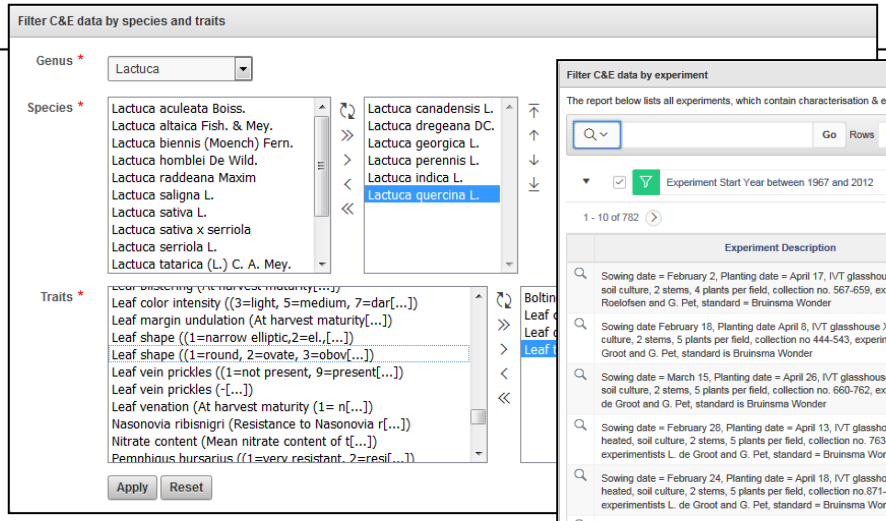


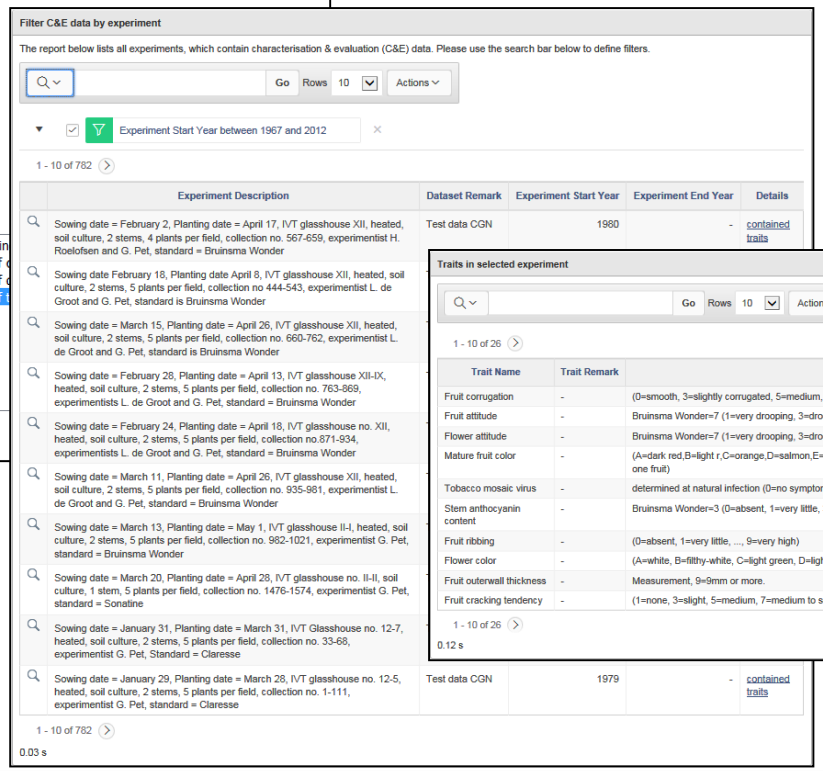


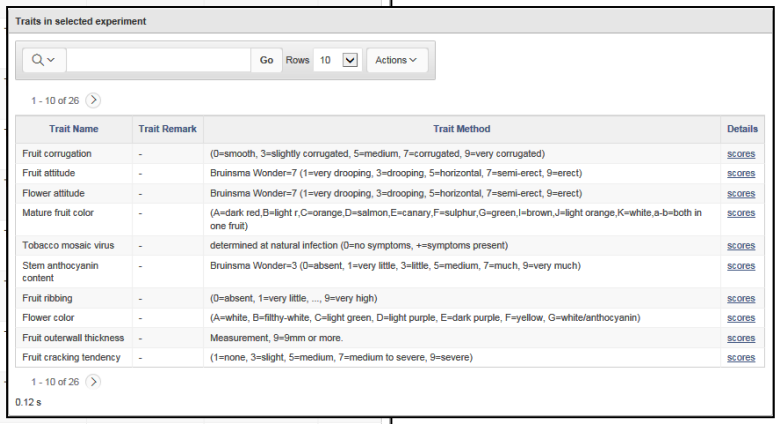
as of 2019-11-25

Phenotypic data search in EURISCO





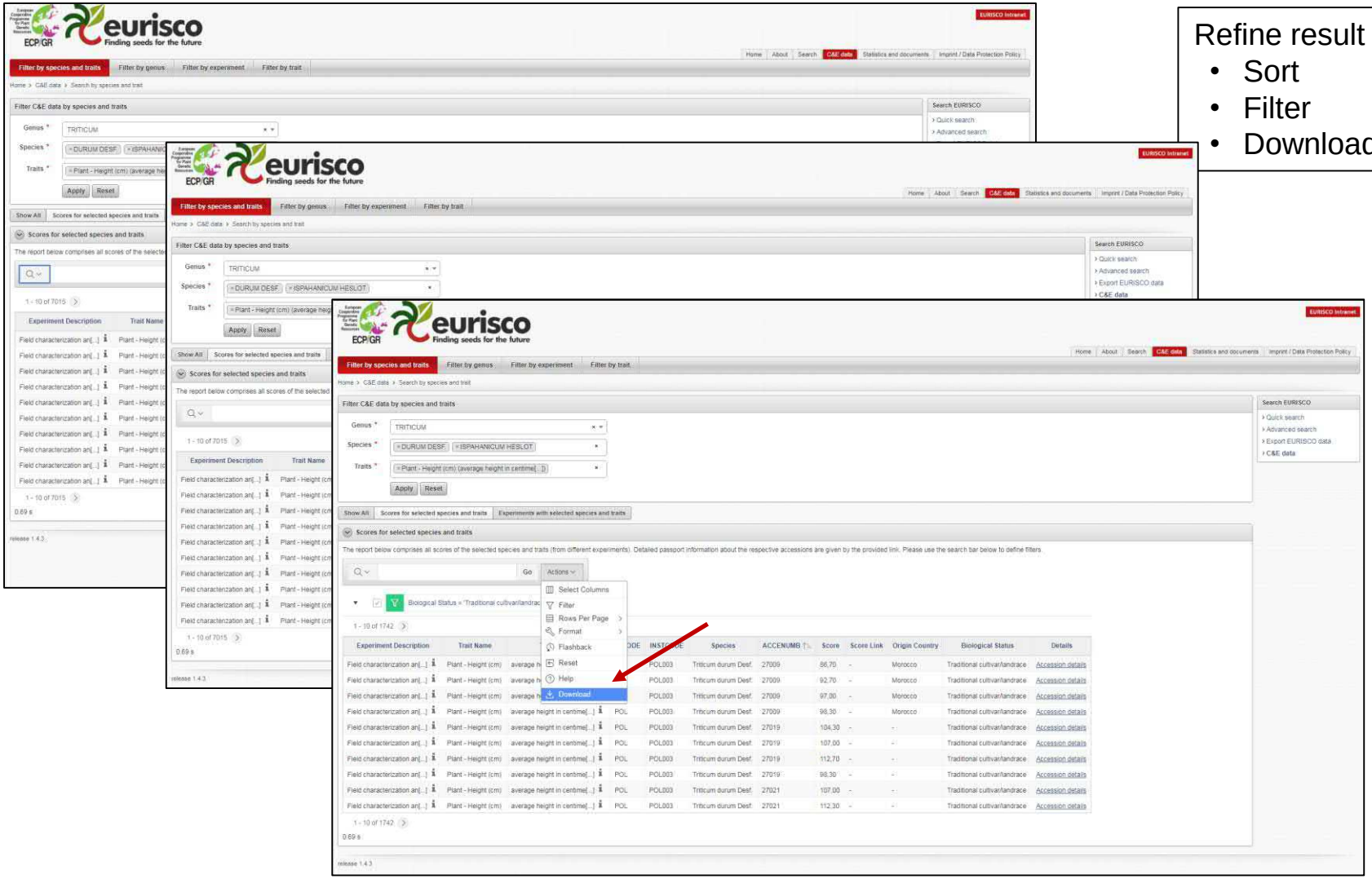




Wizard-based searches for

- Genus
- Species and trait
- Experiment
- Trait

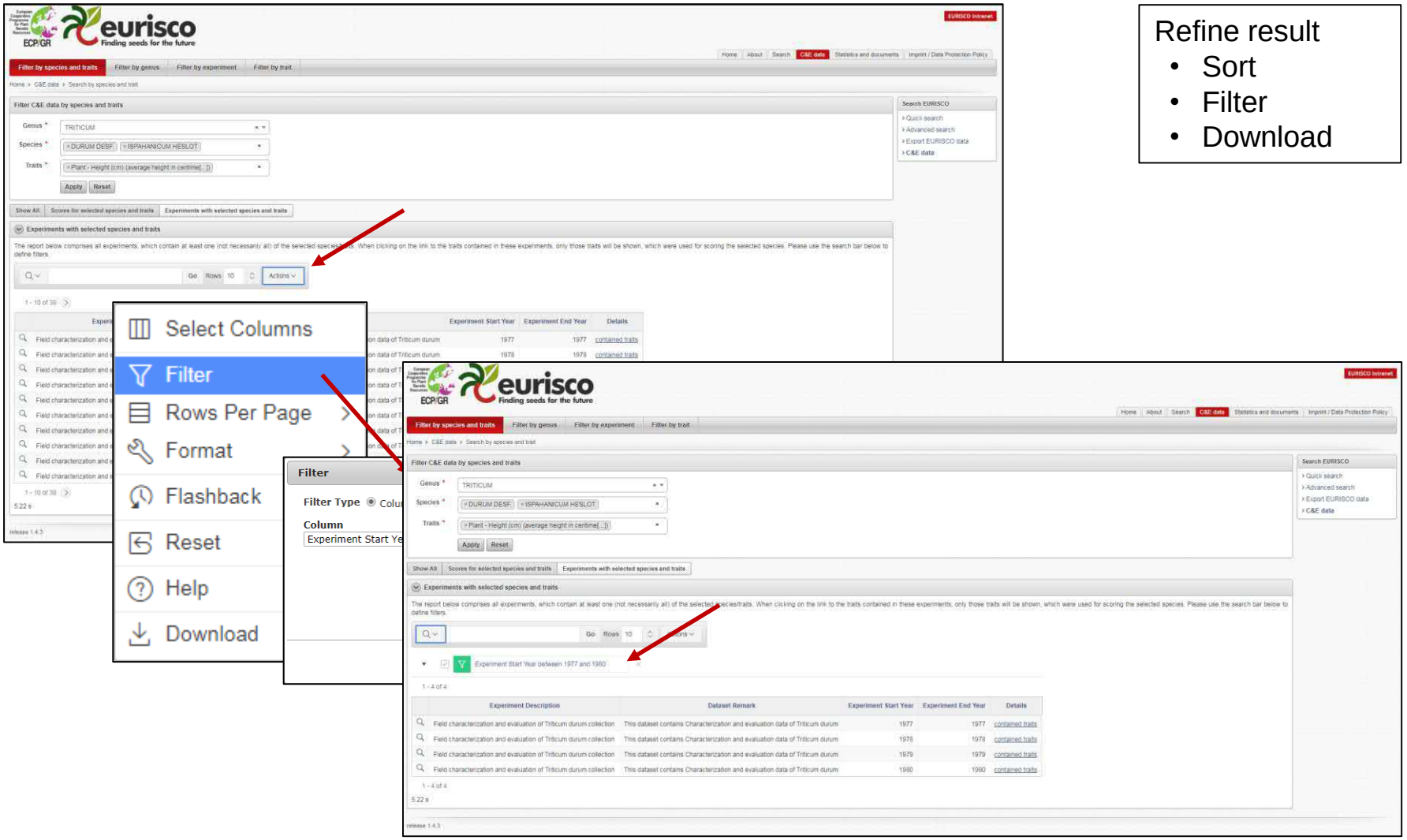
Example I – report of values



Refine result

- Sort
- Filter
- Download

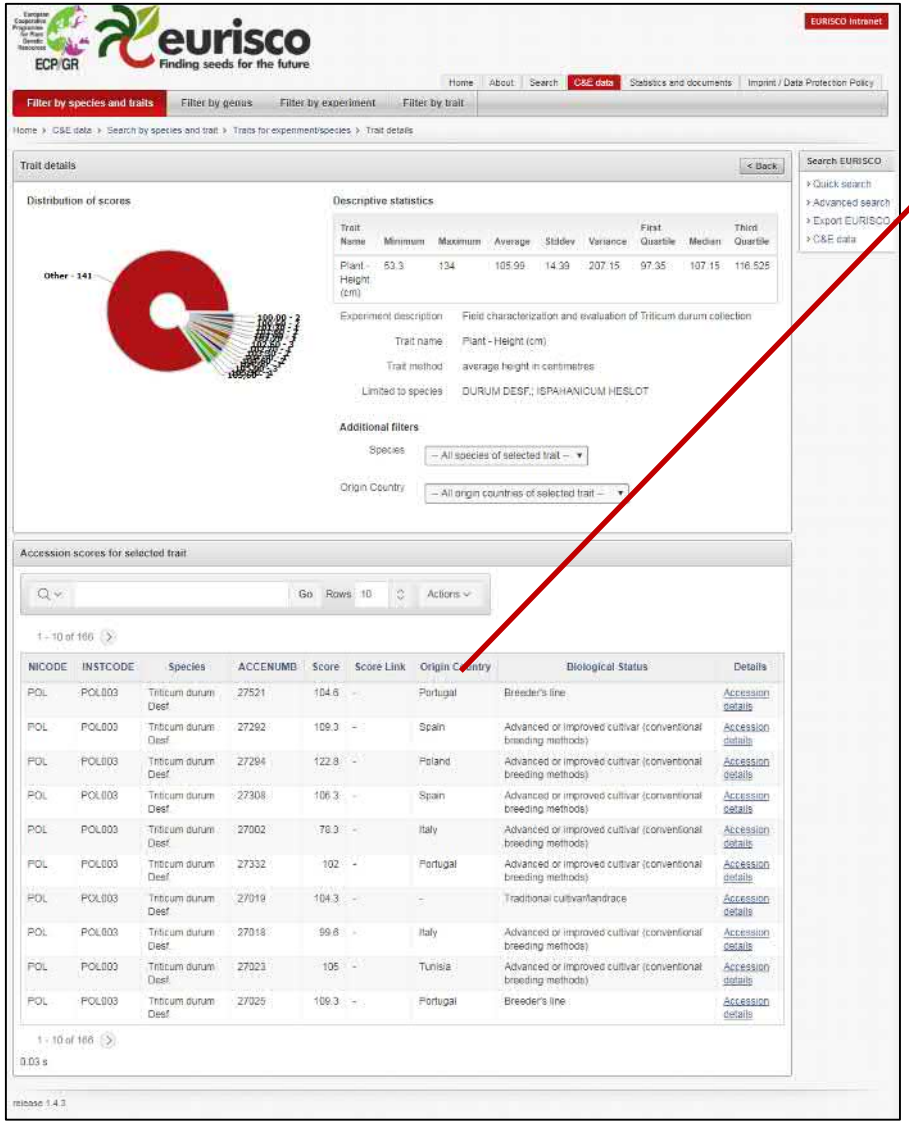
Example I – report of experiments



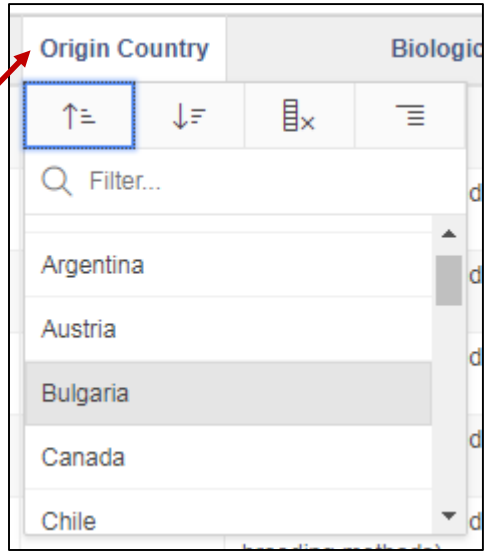
Refine result

- Sort
- Filter
- Download

Example I – scores

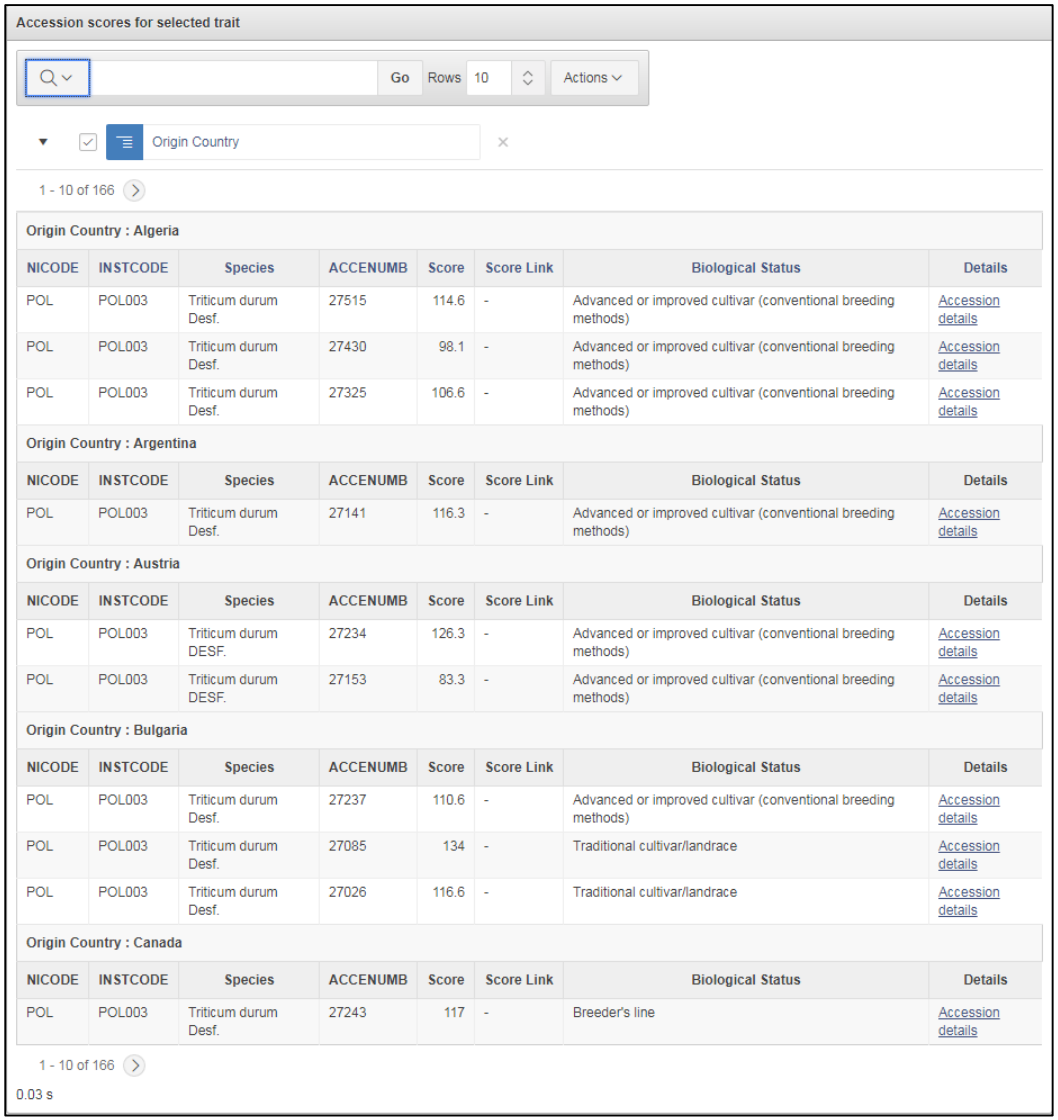


Descriptive statistics



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

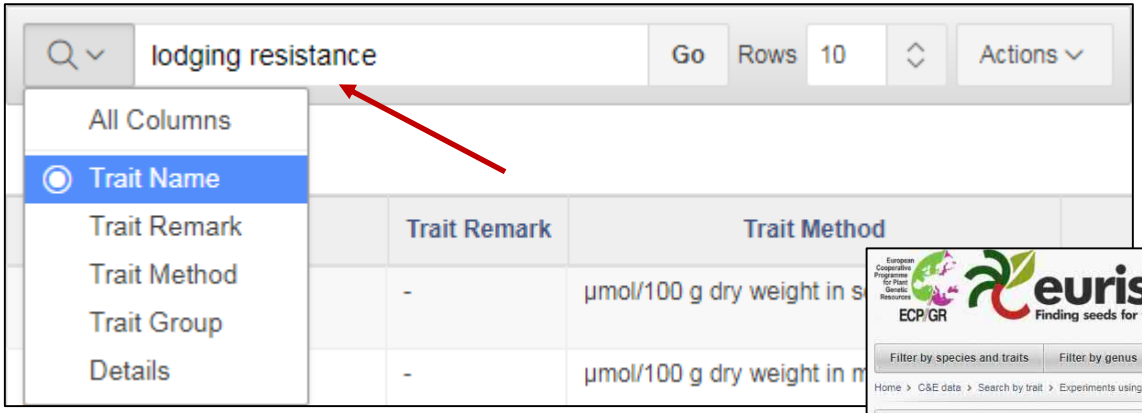
NICODE	INSTCODE	Species	ACCENUMB	Score	Score Link	Biological Status	Details
POL	POL003	Triticum durum Desf.	27243	117	-	Breeder's line	Accession details

1 - 10 of 166

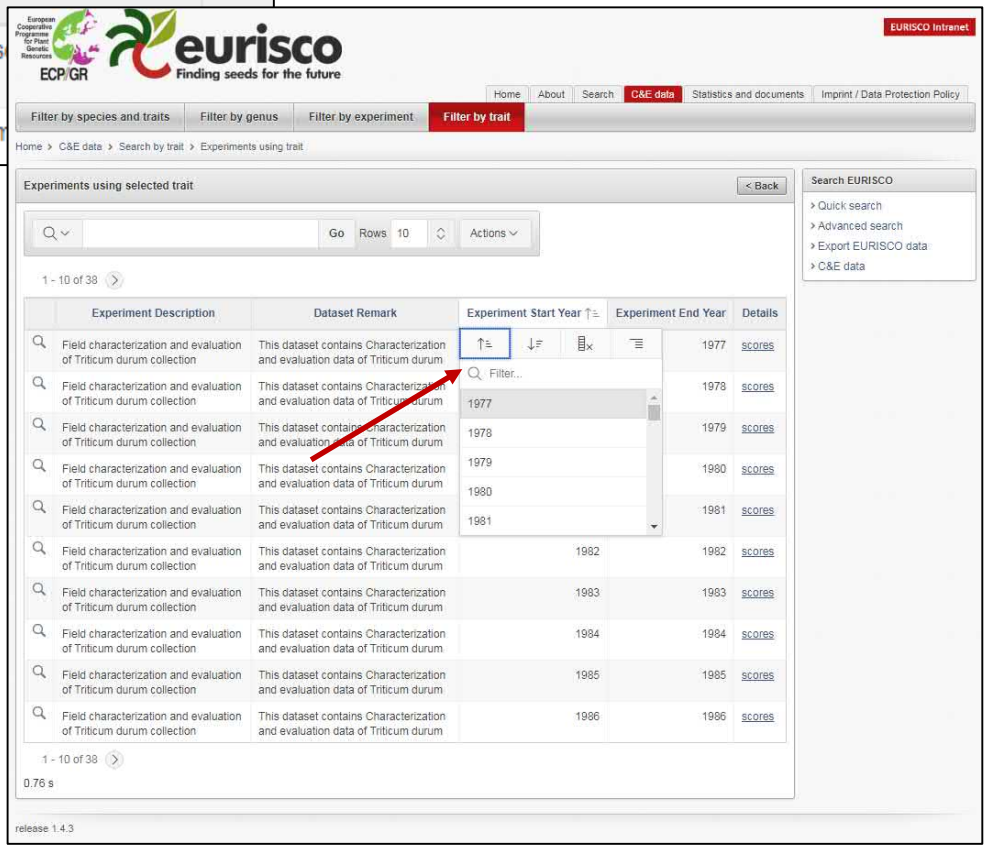
0.03 s

Group values

Example II – trait selection

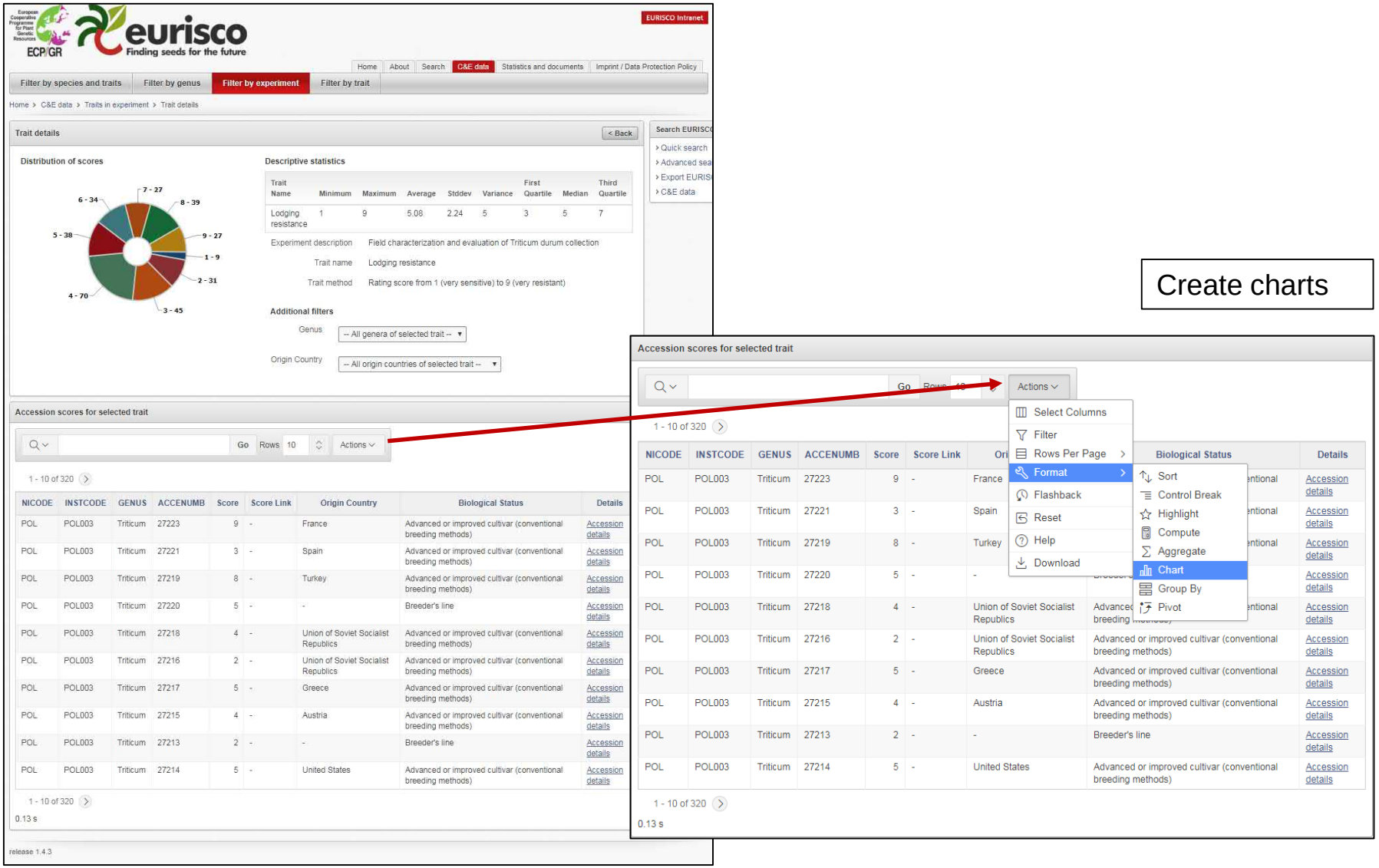


Search for trait of interest



Filter experiments containing the selected trait

Example II – scores



Create charts

Example II – scores

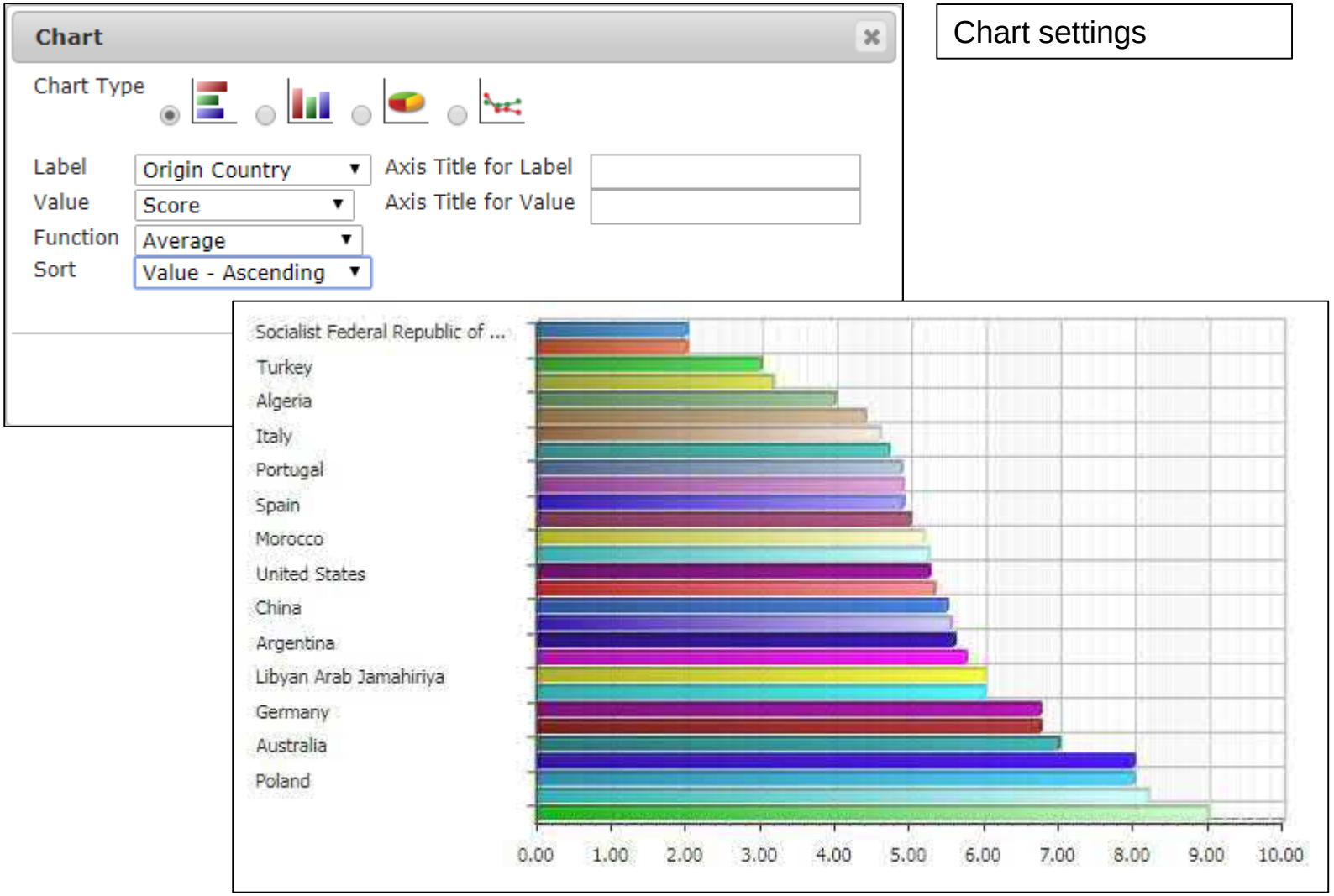


Chart settings

Phenotypic data search in EURISCO

Download experiments

The report below provides the possibility to download whole experiments as MS Excel files, which contain the experiment description, the trait definitions as well as the phenotypic scores. Please use the search bar below to define filters.

Go | Actions

Dataset = 'Characterisation data (1946 - 2012) of barley accessions from DEU146'

1 - 10 of 65

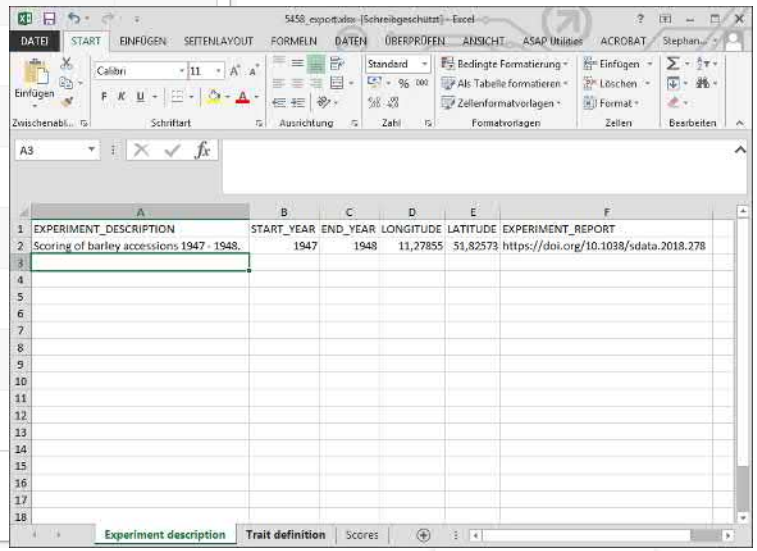
Dataset	Experiment	Start year	End year	File created	Filesize (KB)	
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1945 - 1946.	1945	1946	2019-11-20	152.51	download
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1946 - 1947.	1946	1947	2019-11-20	64.22	download
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1947 - 1948.	1947	1948	2019-11-20	302.51	download
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1948 - 1949.	1948	1949	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1949 - 1950.	1949	1950	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1950 - 1951.	1950	1951	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1951 - 1952.	1951	1952	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1952 - 1953.	1952	1953	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1953 - 1954.	1953	1954	2019-11-20		
Characterisation data (1946 - 2012) of barley accessions from DEU146	Scoring of barley accessions 1954 - 1955.	1954	1955	2019-11-20		

1 - 10 of 65

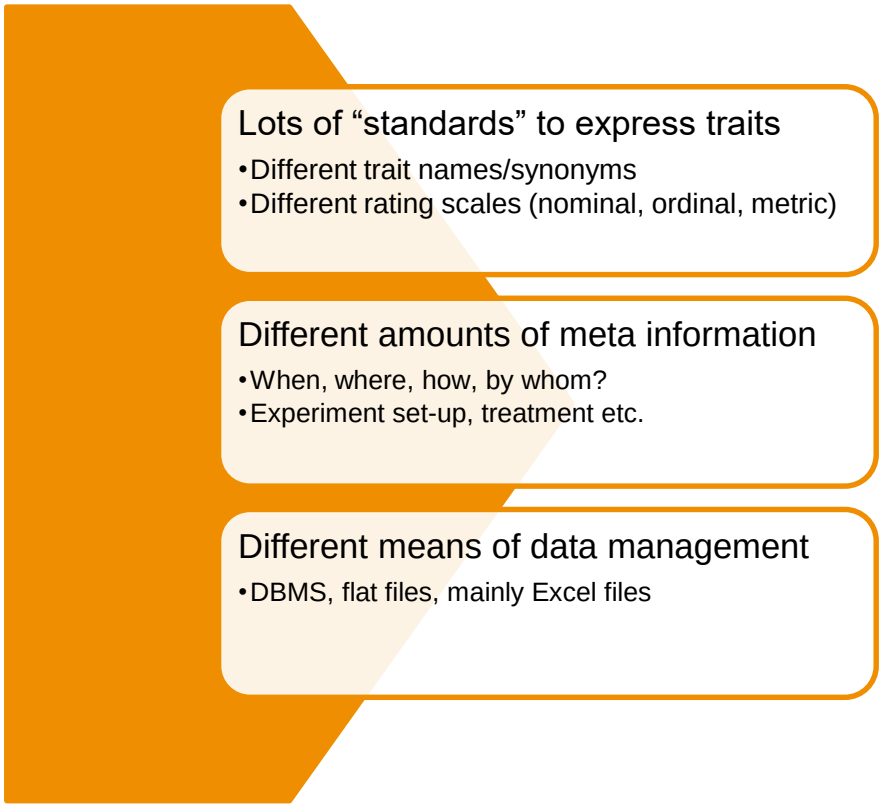
0.02 s

Download of full experiments

- Experiment description
- Trait definition
- Scores



The challenge: Diversity of data



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

→ Input of crop WGs needed

COLLABORATION

Participation in project consortia

- Various ECPGR Grant Scheme Activities

- EUCLEG (Horizon 2020), 2017–2021
  - Coordination: INRA, France
  - Aim: reduction of protein import dependencies of both European and Chinese partners
  - Leader of work package for data management
    - Find data gaps in EURISCO (and try to close them)
    - Manage project data (passport, phenotypic, genetic)

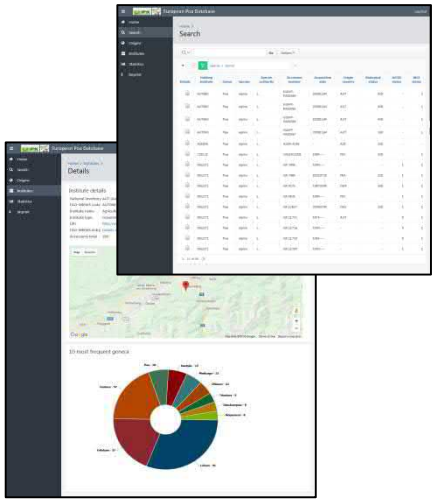
- Farmer's Pride (Horizon 2020), 2018–2020
  - Coordination: University of Birmingham, U.K.
  - Aim: Development of network of *in situ* sites and stakeholders
  - Task leader: Preparation of a concept to extend EURISCO for *in situ* data

- GenRes Bridge (Horizon 2020), (2019–2021)
  - Coordination: European Forest Institute
  - Aim: Join forces of plant, forest and animal genetic resources
  - User training + information system linking

- ECPGR European Evaluation Network (initial funding BLE), 2019–2022
  - Coordination: ECPGR
  - Aim: Implementation of the evaluation network on wheat/barly and vegetable crops
  - Development of infrastructure

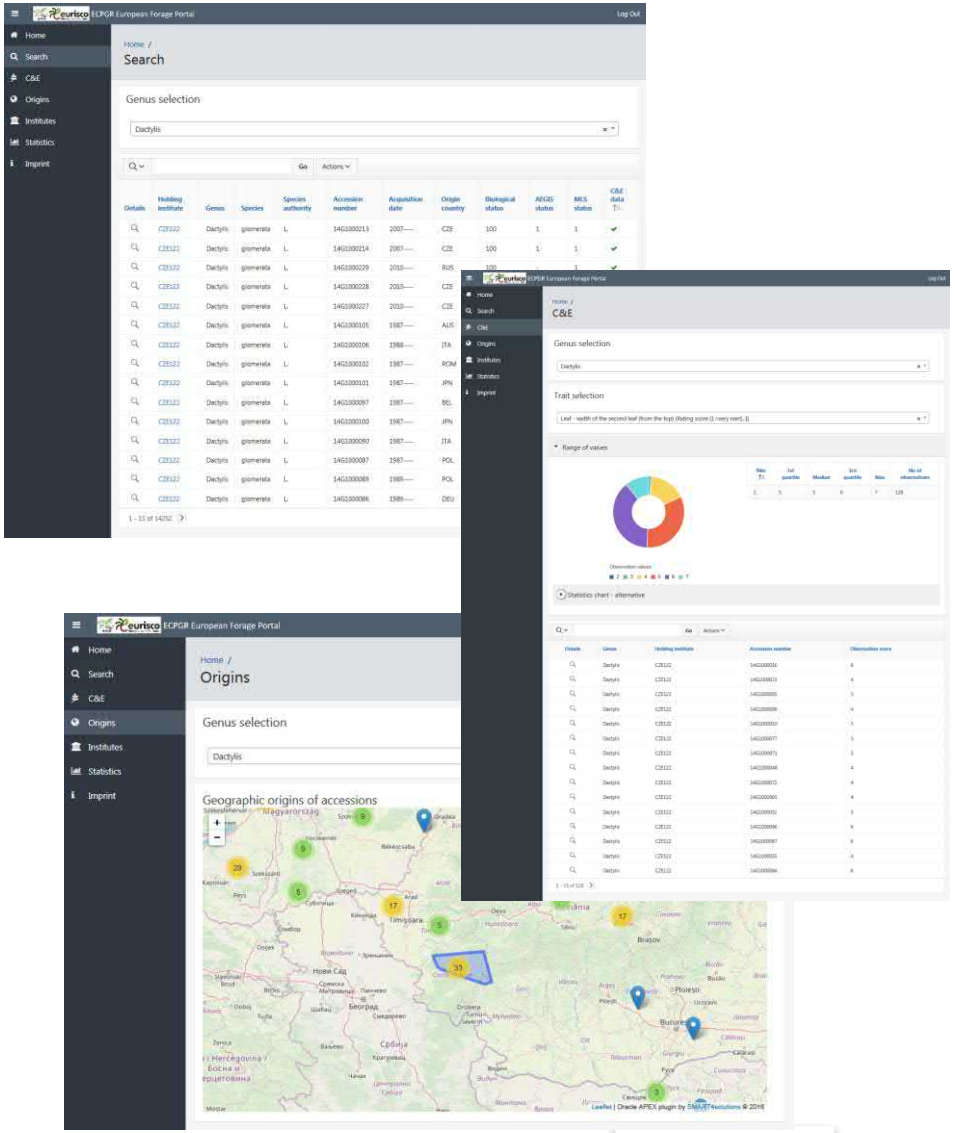
Collaborations

- Genesys (Crop Trust)
  - Synchronisation of passport data
  - European hub for Genesys
- FAO-WIEWS
  - Synchronisation of passport data (non-regular)
- Germinate (JHI)
  - Close interlinking
- GLIS (ITPGRFA)
  - EURISCO provides a service for registering accessions for DOIs
- ECPGR crop working groups
  - Backend support for crop portals



Support of ECPGR Central Crop Databases

- Cooperation with various crop WGs
- Example: European Forage Portal (Forage WG)
  - Browse
    - Passport data
    - Phenotypic data
  - Selection via map
  - Statistics
- Automatically updated from EURISCO
- Blueprint for other crop portals



M. Grau / IPK

THANK YOU FOR YOUR ATTENTION