# European Plant Genetic Resources in the EURISCO information system

First Meeting of the ECPGR Berries Working Group, 14<sup>th</sup> to 16<sup>th</sup> January 2020, Dresden, Germany

Stephan Weise 15 January 2020



# 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

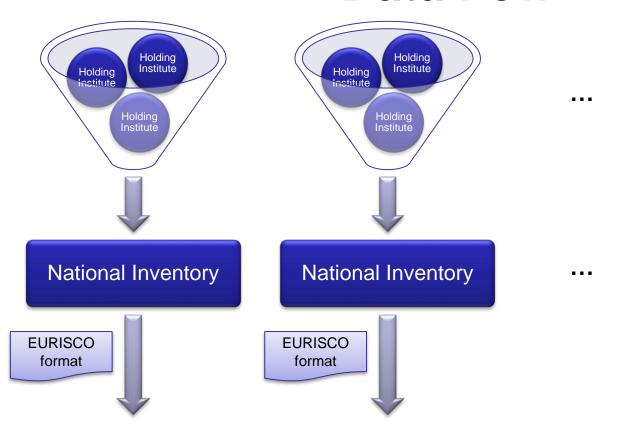


https://upload.wikimedia.org/wikipedia/commons/8/81/Europe\_countries\_map\_2.png





#### **Data flow**



398 germplasm collections

43 member countries

#### **EURISCO**

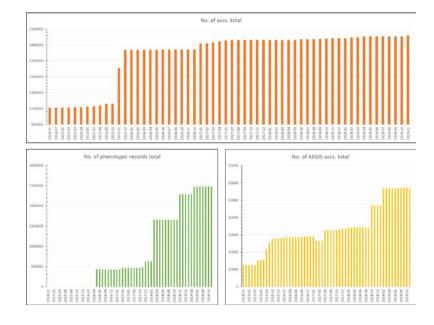
ex situ PGR data





#### **Contents of EURISCO**

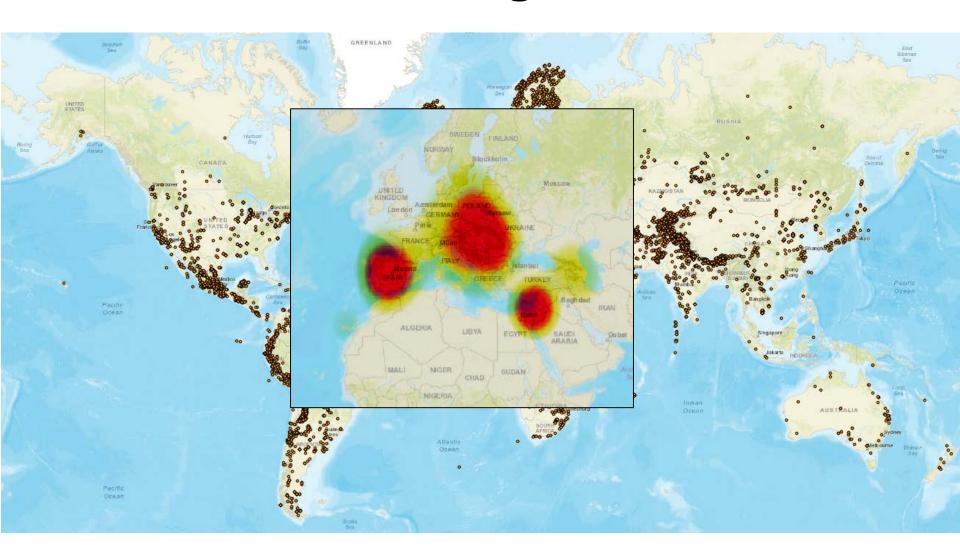
- 2,019,698 accessions
- 6,392 genera (including synonyms, spelling variants)
- 43,220 species names
- 444,360 MLS accessions
- 56,934 AEGIS accessions
- 60,500 DOIs







# **Collecting sites**







# **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

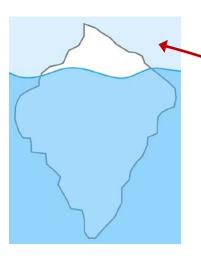




### **EURISCO WEB**







#### 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<sup>1,\*</sup>, Markus Oppermann<sup>1</sup>, Lorenzo Maggioni<sup>2</sup>, Theo van Hintum<sup>3</sup> and Helmut Knüpffer<sup>1</sup>

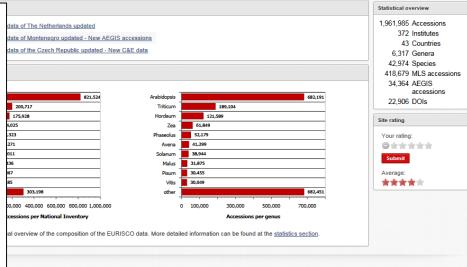
<sup>1</sup>Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Stadt Seeland, Germany, <sup>2</sup>Bioversity International, Via dei Tre Denari 472/a, 00057 Maccarese (Fiumicino), Rome, Italy and <sup>3</sup>Centre for Genetic Resources, The Netherlands (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

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.



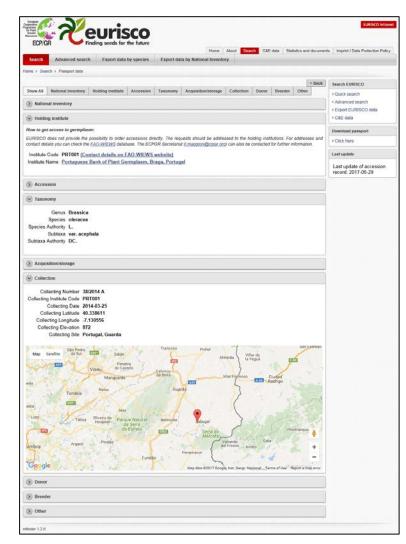
53 (sub)versions since 2014





### Passport data search in EURISCO

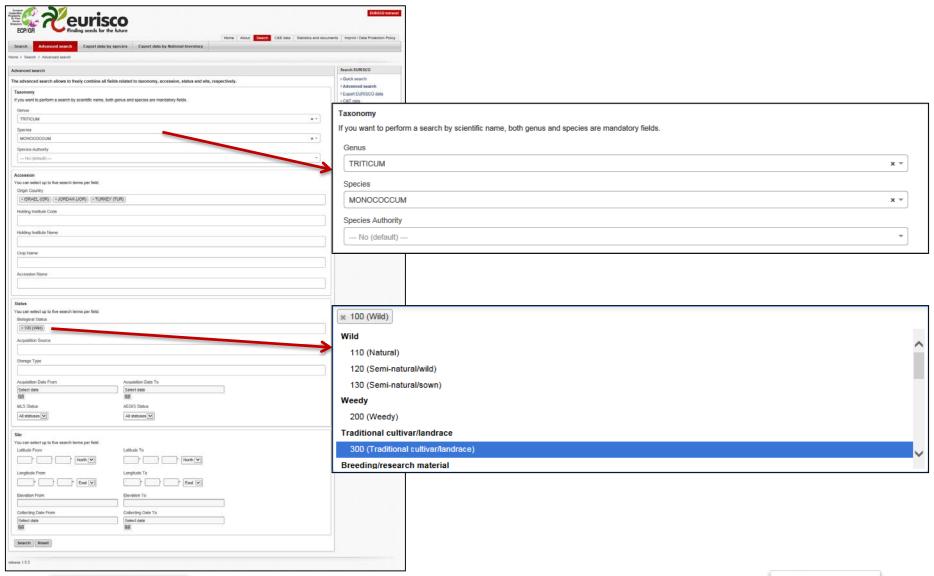
- Four standard searches:
  - Taxonomy (incl. synonyms)
  - Accession
  - Biological status
  - Collecting site
- Advanced search
- Different user-specific export features







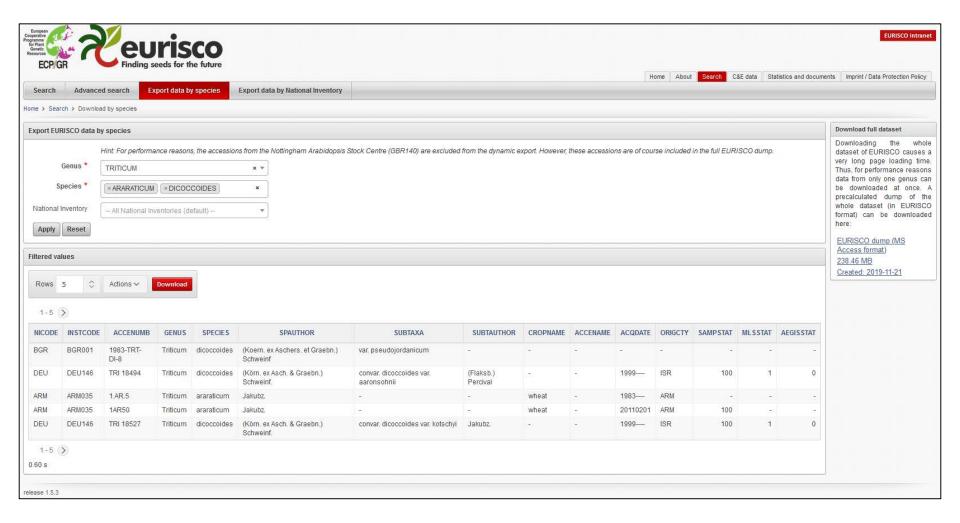
#### Advanced search form







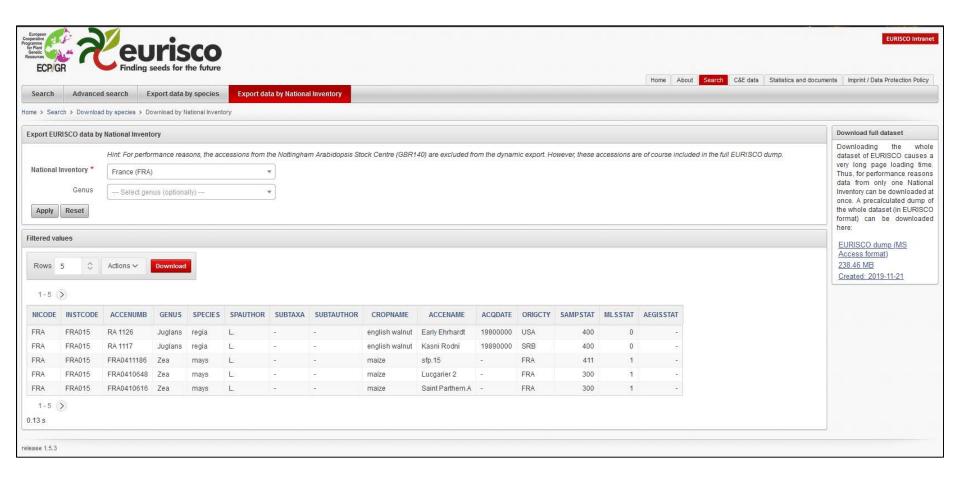
# User specific export – by species







# 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!





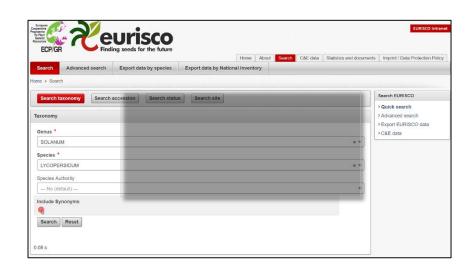
## Taxonomy search

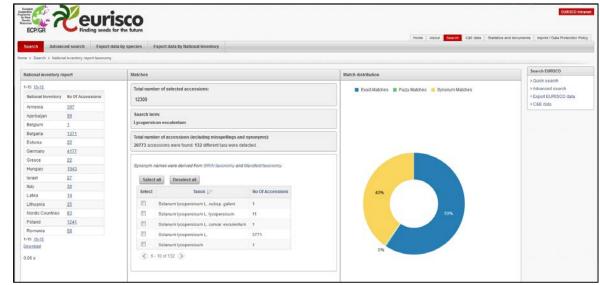
- Rework of the search interface
  - Allows to include synonyms
  - Shows the distribution of matching types

Significant performance

improvement











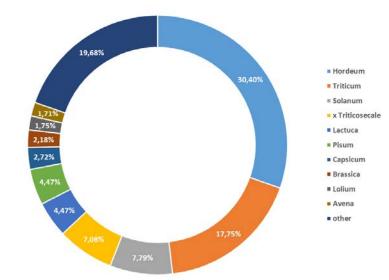
#### PHENOTYPIC DATA

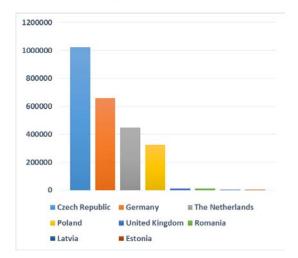




### Phenotypic data

- 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



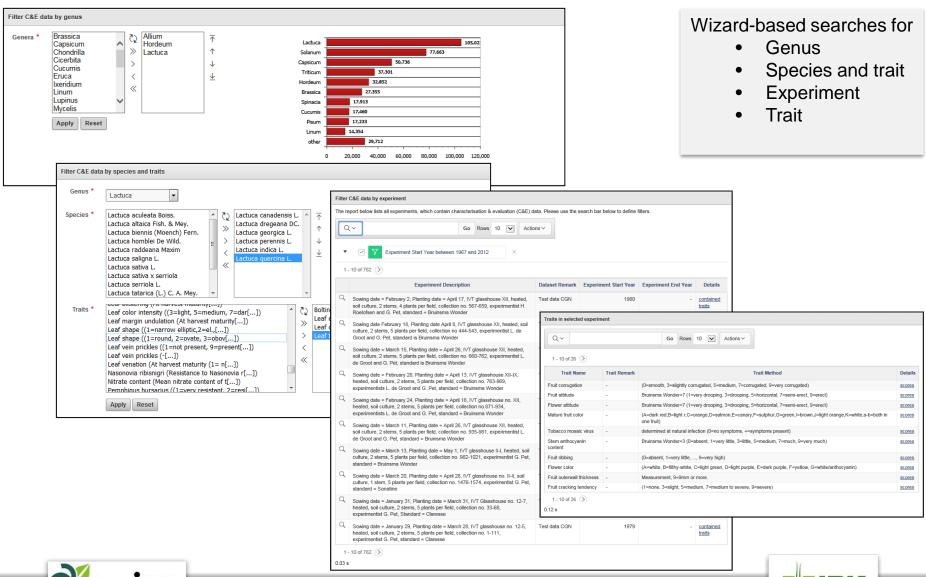






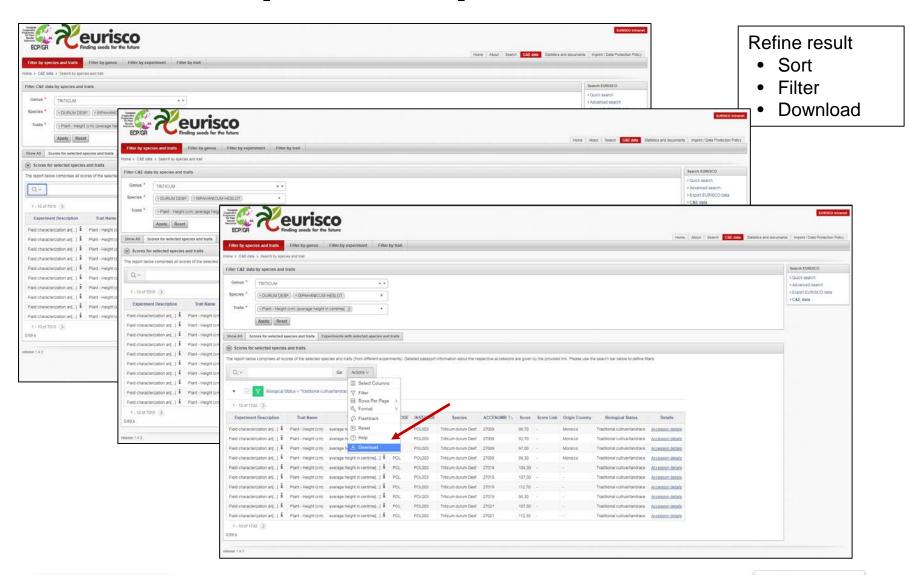


#### Phenotypic data search in EURISCO





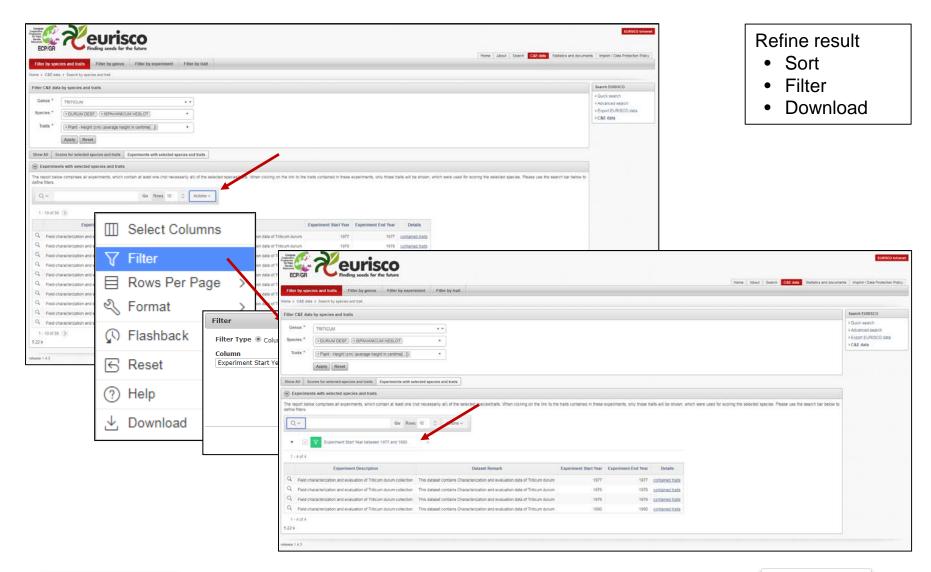
### Example I – report of values







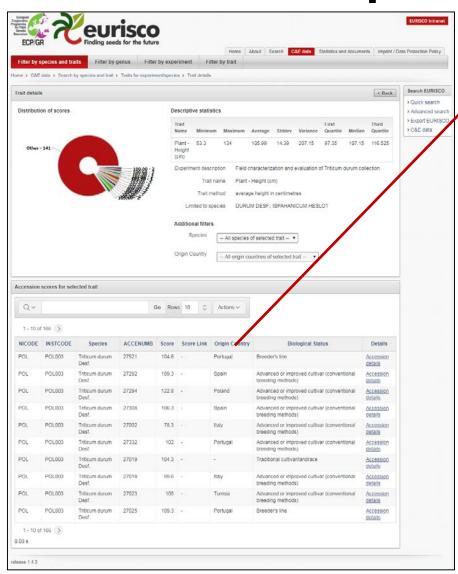
### Example I – report of experiments



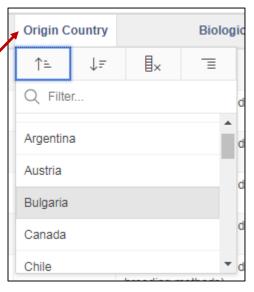




#### Example I – scores



Descriptive statistics

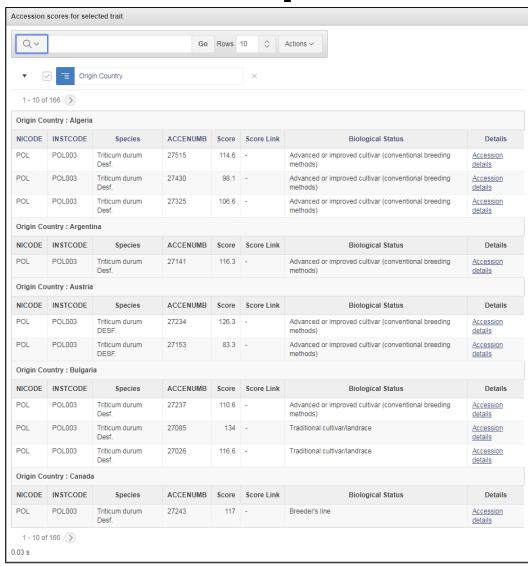


Additional filters, e.g. origin country





#### Example I – scores

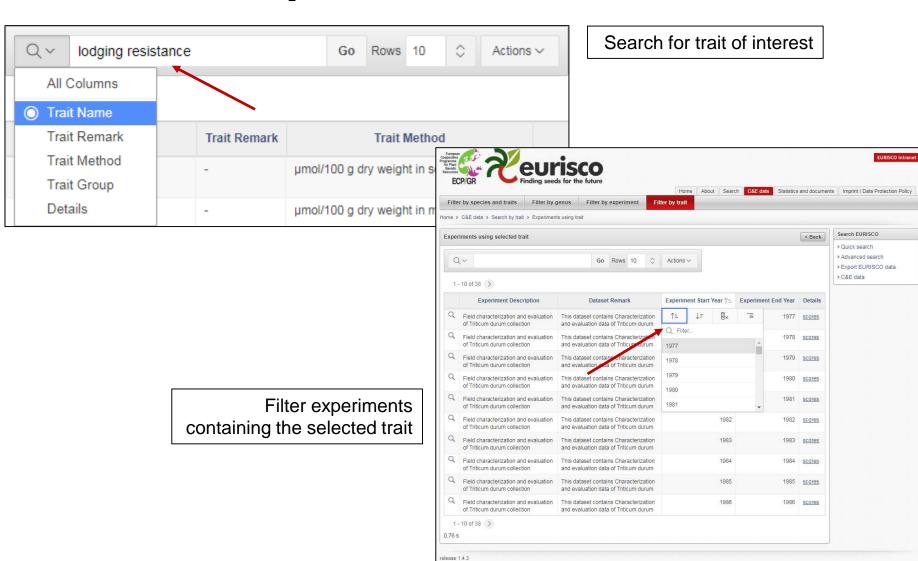


Group values





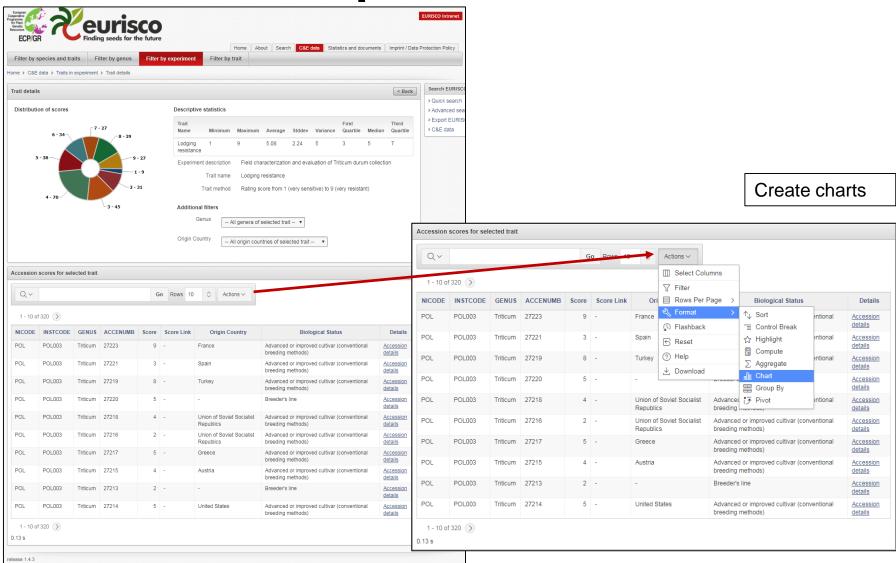
### Example II – trait selection







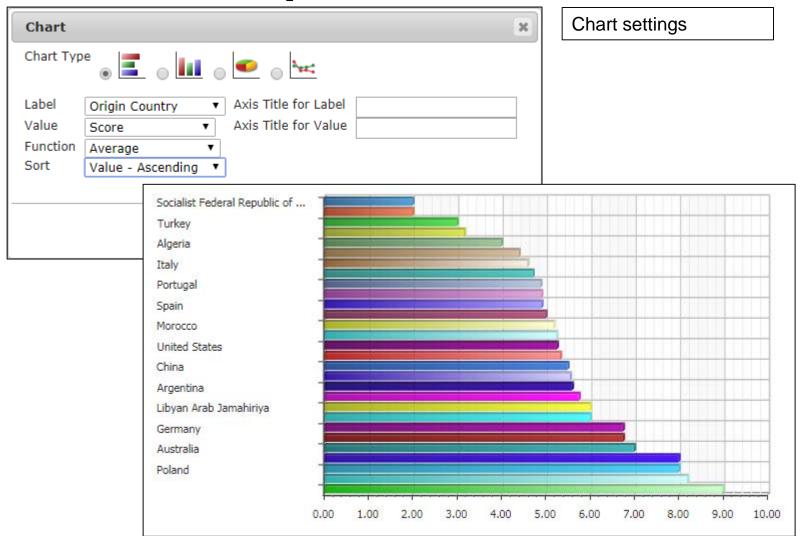
# Example II – scores







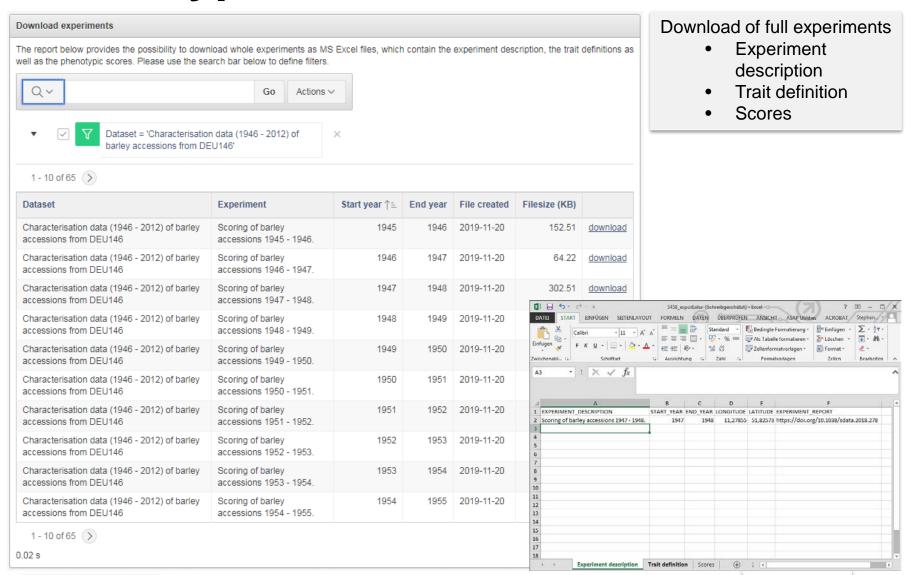
#### Example II – scores







### Phenotypic data search in EURISCO







## The challenge: Diversity of data

#### Lots of "standards" to express traits

- Different trait names/synonyms
- Different rating scales (nominal, ordinal, metric)

#### Different amounts of meta information

- •When, where, how, by whom?
- Experiment set-up, treatment etc.

#### Different means of data management

•DBMS, flat files, mainly Excel files





### **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
- AGENT (Horizon 2020), (2020–2025)
  - Coordination: IPK Gatersleben
  - Aim: Activate genebanks and facilitate access to plant genetic resources
  - Task leader: Data exchange and representation; infrastructure for managing and analysing genotypic and phenotypic data about genetic resources









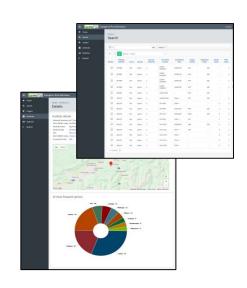






#### **Collaborations**

- 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

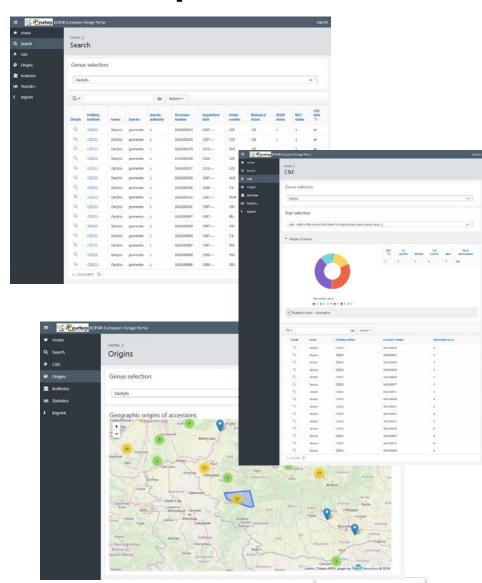






#### **Support of ECPGR Central Crop Databases**

- 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



