

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

Annual meeting of the European Evaluation Network (EVA) for lettuce,
30th March 2020, online via MS Teams

Stephan Weise
30 March 2020



OVERVIEW

Background

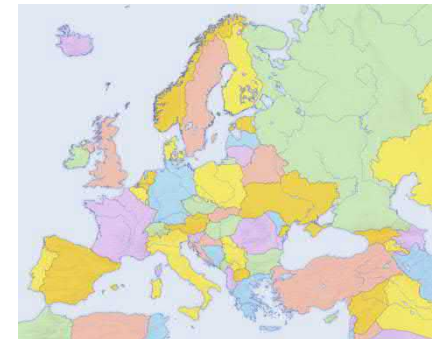
- What is EURISCO?
 - European information system for plant genetic resources
 - Search catalogue for *ex situ* collections
 - Accession-level information system
- Purpose
 - Provides passport data and phenotypic data about plant germplasm accessions maintained in Europe
 - Assists in meeting national 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

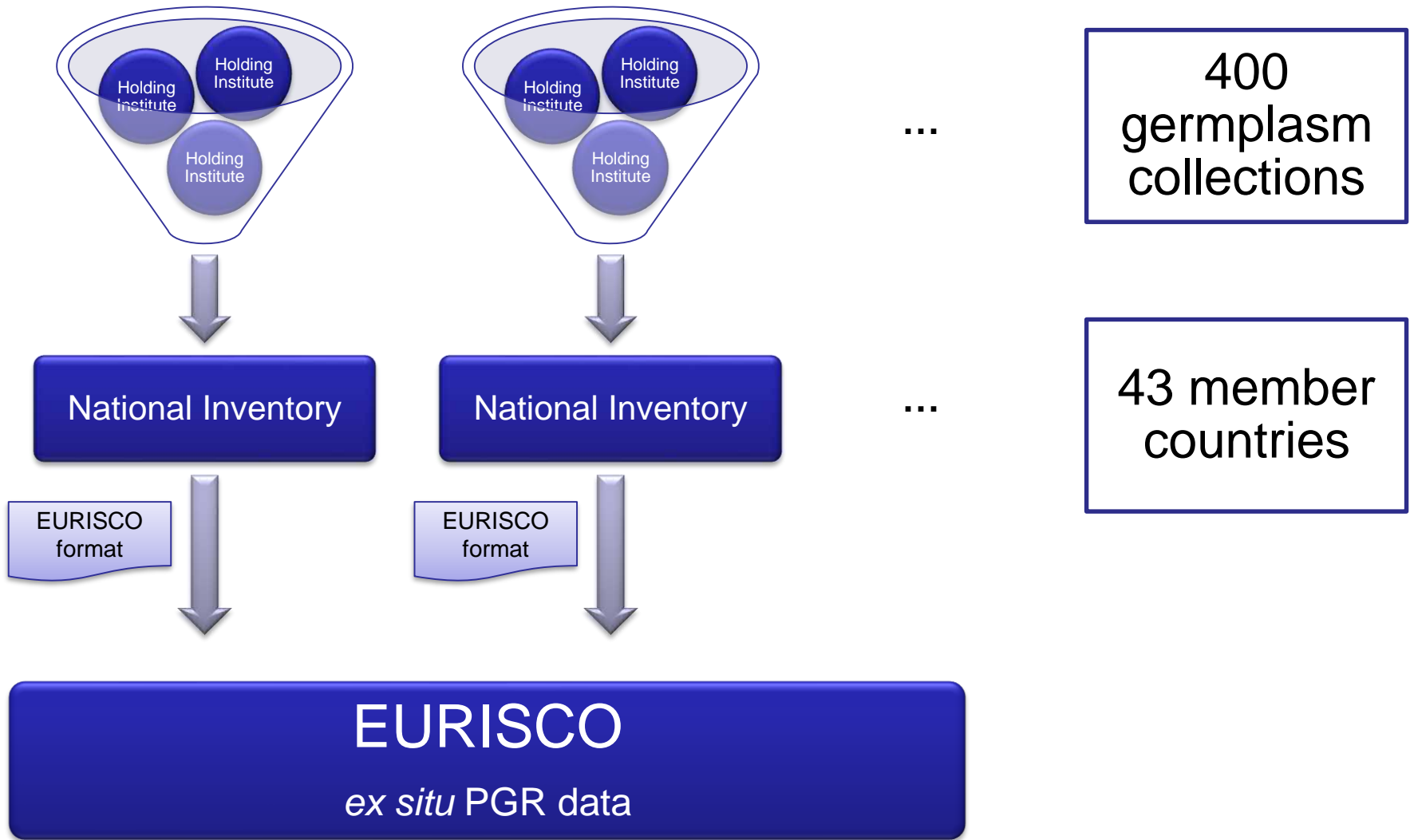
Development

- 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



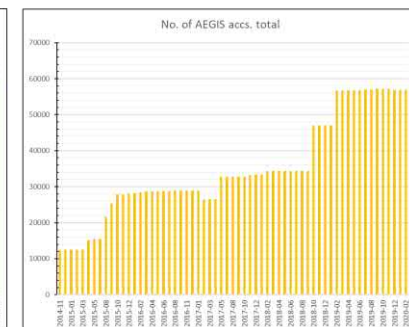
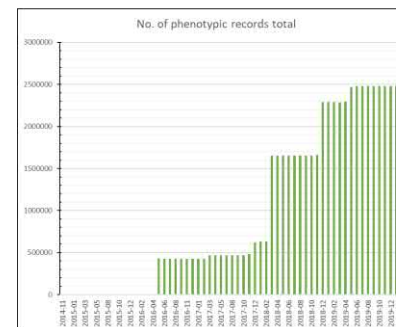
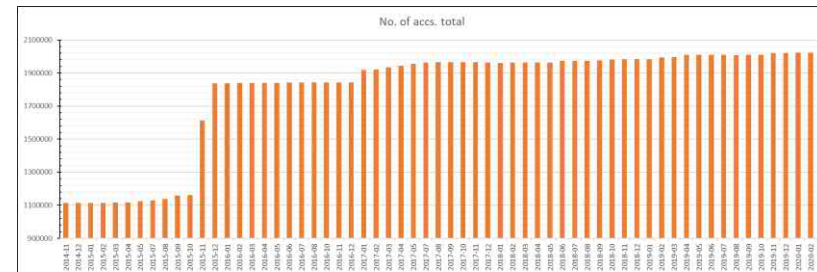
https://upload.wikimedia.org/wikipedia/commons/8/81/Europe_countries_map_2.png

Data flow



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 Oppermann¹, Lorenzo Maggioni², Theo van Hintum³ and Helmut Knüpffer¹

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Stadt Seeland, Germany, ²Biodiversity International, Via dei Tre Denari 472/a, 00057 Maccarese (Fiumicino), Rome, Italy and ³Centre 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

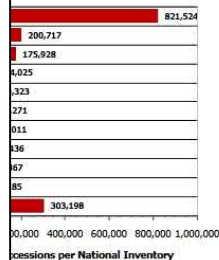
The European Search Catalogue for Plant Genetic Resources, EURISCO, provides information about 1.8 million crop plant accessions preserved by almost 400 institutes in Europe and beyond. EURISCO is being maintained on behalf of the European Cooperative Programme for Plant Genetic Resources. It is based on a network of National Inventories of 42

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

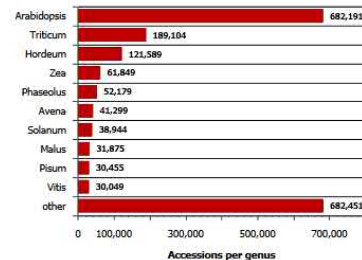
data of The Netherlands updated

data of Montenegro updated - New AEGIS accessions

data of the Czech Republic updated - New C&E data



Statistical overview of the composition of the EURISCO data. More detailed information can be found at the [statistics section](#).



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

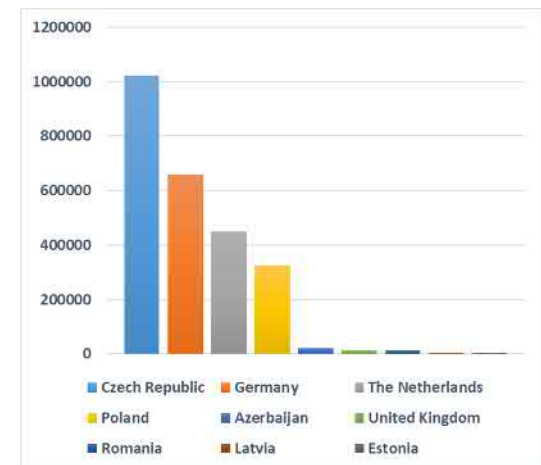
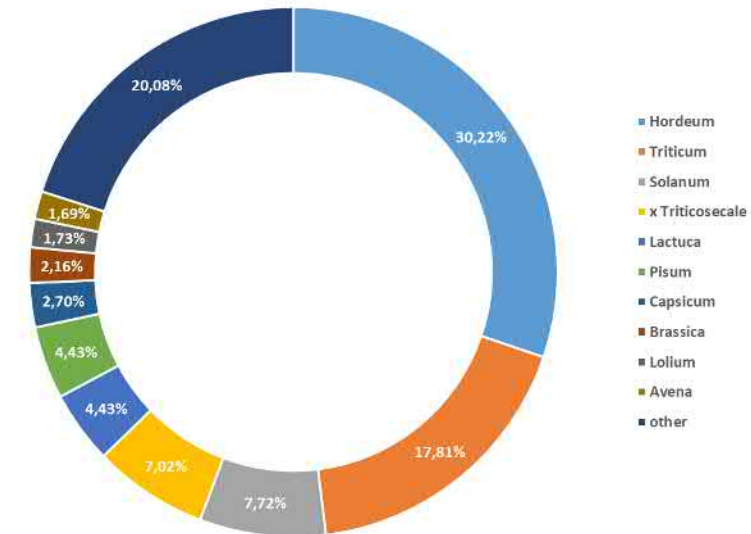
The screenshot displays the EURISCO web interface for a passport data search. The header includes the EURISCO logo and navigation links. The main content area is divided into several sections:

- National inventory:** Shows the accession code PRI001 and the holding institution name, Portuguese Bank of Plant Germplasm, Braga, Portugal.
- Taxonomy:** Lists the taxonomic classification: Genus *Brassica*, Species *oleracea*, Species Authority L., Subtaxa *var. acephala*, and Subtaxa Authority DC.
- Acquisition/storage:** Provides collection details: Collecting Number 382014 A, Collecting Institute Code PRI001, Collecting Date 2014-03-29, Collecting Latitude 40.338611, Collecting Longitude -7.130956, Collecting Elevation 872, and Collecting Site Portugal, Guarda.
- Map:** A Google Maps view showing the collection site in Guarda, Portugal, with a red pin and a map scale of 1:2.5.
- Donor, Breeder, and Other:** Sections for additional information, currently empty.

The right sidebar contains search options and a last update record of 2017-05-29.

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

Filter C&E data by genus

Genera *

- Brassica
- Capsicum
- Chondrilla
- Cicerbita
- Cucumis
- Eruca
- Ixeridium
- Linum
- Lupinus
- Mycelis

Apply Reset

Alium
Hordeum
Lactuca

Genus	Count
Lactuca	105,020
Solanum	77,663
Capsicum	50,736
Triticum	37,301
Hordeum	32,852
Brassica	27,355
Spinacia	17,913
Cucumis	17,460
Pisum	17,233
Linum	14,354
other	29,712

Wizard-based searches for

- Genus
- Species and trait
- Experiment
- Trait

Filter C&E data by species and traits

Genus * Lactuca

Species *

- Lactuca aculeata Boiss.
- Lactuca altaica Fish. & Mey.
- Lactuca biennis (Moench) Fern.
- Lactuca homblei De Wild.
- Lactuca raddeana Maxim
- Lactuca saligna L.
- Lactuca sativa L.
- Lactuca sativa x serriola
- Lactuca serriola L.
- Lactuca tatarica (L.) C. A. Mey.
- Lactuca canadensis L.
- Lactuca dregeana DC.
- Lactuca georgica L.
- Lactuca perennis L.
- Lactuca indica L.
- Lactuca quercina L.

Traits *

- Leaf color intensity ((3=light, 5=medium, 7=dar[...]))
- Leaf margin undulation (At harvest maturity[...])
- Leaf shape ((1=narrow elliptic, 2=el., [...]))
- Leaf shape ((1=round, 2=ovate, 3=obov[...]))
- Leaf vein prickles ((1=not present, 9=present[...]))
- Leaf vein prickles (-[...])
- Leaf venation (At harvest maturity (1 = n[...]))
- Nasonovia ribisnigri (Resistance to Nasonovia r[...])
- Nitrate content (Mean nitrate content of t[...])
- Pemphigus hirsarius ((1=very resistant, 2=resil[...]))

Apply Reset

Filter C&E data by experiment

The report below lists all experiments, which contain characterisation & evaluation (C&E) data. Please use the search bar below to define filters.

Experiment Start Year between 1967 and 2012

1 - 10 of 782

Experiment Description	Dataset Remark	Experiment Start Year	Experiment End Year	Details
Sowing date = February 2, Planting date = April 17, IVT glasshouse XII, heated, soil culture, 2 stems, 4 plants per field, collection no. 567-659, experimentalist H. Roelofsan and G. Pet, standard = Bruinsma Wonder	Test data CGN	1980	-	contained traits
Sowing date February 18, Planting date April 8, IVT glasshouse XII, heated, soil culture, 2 stems, 5 plants per field, collection no. 444-543, experimentalist L. de Groot and G. Pet, standard = Bruinsma Wonder				
Sowing date = March 15, Planting date = April 26, IVT glasshouse XII, heated, soil culture, 2 stems, 5 plants per field, collection no. 660-762, experimentalist L. de Groot and G. Pet, standard = Bruinsma Wonder				
Sowing date = February 28, Planting date = April 13, IVT glasshouse XII-IX, heated, soil culture, 2 stems, 5 plants per field, collection no. 935-981, experimentalist L. de Groot and G. Pet, standard = Bruinsma Wonder				
Sowing date = February 24, Planting date = April 18, IVT glasshouse no. XII, heated, soil culture, 2 stems, 5 plants per field, collection no. 871-934, experimentalists L. de Groot and G. Pet, standard = Bruinsma Wonder				
Sowing date = March 11, Planting date = April 26, IVT glasshouse XII, heated, soil culture, 2 stems, 5 plants per field, collection no. 982-1021, experimentalist G. Pet, standard = Bruinsma Wonder				
Sowing date = March 13, Planting date = May 1, IVT glasshouse II-I, heated, soil culture, 2 stems, 5 plants per field, collection no. 982-1021, experimentalist G. Pet, standard = Bruinsma Wonder				
Sowing date = March 20, Planting date = April 28, IVT glasshouse no. II-II, soil culture, 1 stem, 5 plants per field, collection no. 1476-1574, experimentalist G. Pet, standard = Sonatine				
Sowing date = January 31, Planting date = March 31, IVT Glasshouse no. 12-7, heated, soil culture, 2 stems, 5 plants per field, collection no. 33-65, experimentalist G. Pet, Standard = Claessee				
Sowing date = January 29, Planting date = March 28, IVT glasshouse no. 12-5, heated, soil culture, 2 stems, 5 plants per field, collection no. 1-111, experimentalist G. Pet, standard = Claessee	Test data CGN	1979	-	contained traits

1 - 10 of 782

0.03 s

Traits in selected experiment

1 - 10 of 26

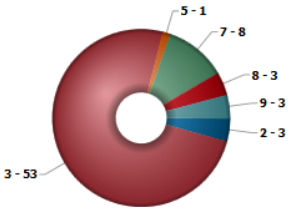
Trait Name	Trait Remark	Trait Method	Details
Fruit corrugation	-	(0=smooth, 3=slightly corrugated, 5=medium, 7=corrugated, 9=very corrugated)	scores
Fruit attitude	-	Bruinsma Wonder=7 (1=very drooping, 3=drooping, 5=horizontal, 7=semi-erect, 9=erect)	scores
Flower attitude	-	Bruinsma Wonders=7 (1=very drooping, 3=drooping, 5=horizontal, 7=semi-erect, 9=erect)	scores
Mature fruit color	-	(A=dark red, B=light r, C=orange, D=salmon, E=canary, F=sulphur, G=green, I=brown, J=light orange, K=white, a-b=both in one fruit)	scores
Tobacco mosaic virus	-	determined at natural infection (0=no symptoms, +=symptoms present)	scores
Stem anthocyanin content	-	Bruinsma Wonder=3 (0=absent, 1=very little, 3=little, 5=medium, 7=much, 9=very much)	scores
Fruit ribbing	-	(0=absent, 1=very little, ..., 9=very high)	scores
Flower color	-	(A=white, B=filly-white, C=light green, D=light purple, E=dark purple, F=yellow, G=white/anthocyanin)	scores
Fruit outerwall thickness	-	Measurement, 9=9mm or more.	scores
Fruit cracking tendency	-	(1=none, 3=slight, 5=medium, 7=medium to severe, 9=severe)	scores

1 - 10 of 26

0.12 s

Phenotypic data search in EURISCO

Trait details



Descriptive statistics

Trait Name	Minimum	Maximum	Average	Stddev	Variance	First Quartile	Median	Third Quartile
Fruit attitude	2	9	3.9	1.93	3.72	3	3	3

Experiment description: Sowing date = February 2, Planting date = April 17, IVT glasshouse XII, heated, soil culture, 2 stems, 4 plants per field, collection no. 567-659, experimentist H. Roelofsens and G. Pet, standard = Bruinsma Wonder

Trait name: Fruit attitude

Additional filters

Genus:

Origin Country:

Accession scores for selected trait

Rows 10

1 - 10 of 71 >

NICODE	INSTCODE	GENUS	ACCENUMB	Score	Score Link	Origin Country	Biological Status	Details
NLD	NLD037	Capsicum	CGN16913	3	-	Germany	Advanced or improved cultivar (conventional breeding methods)	Accession details
NLD	NLD037	Capsicum	CGN16914	3	-	Netherlands	Advanced or improved cultivar (conventional breeding methods)	Accession details
NLD	NLD037	Capsicum	CGN16916	8	-	Israel	Advanced or improved cultivar (conventional breeding methods)	Accession details
NLD	NLD037	Capsicum	CGN16917	7	-	Israel	Advanced or improved cultivar (conventional breeding methods)	Accession details
NLD	NLD037	Capsicum	CGN16918	3	-	Hungary	Traditional cultivar/landrace	Accession details
NLD	NLD037	Capsicum	CGN16919	3	-	Hungary	Advanced or improved cultivar (conventional breeding methods)	Accession details
NLD	NLD037	Capsicum	CGN16920	7	-	-	-	Accession details
NLD	NLD037	Capsicum	CGN16904	9	-	-	-	Accession details
NLD	NLD037	Capsicum	CGN16905	3	-	-	Traditional cultivar/landrace	Accession details
NLD	NLD037	Capsicum	CGN16906	3	-	-	Advanced or improved cultivar (conventional breeding methods)	Accession details

1 - 10 of 71 >

0.02 s

- 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

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

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

Q Go Rows 10 Actions

Trait Name contains 'flowering time'

1 - 10 of 60

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