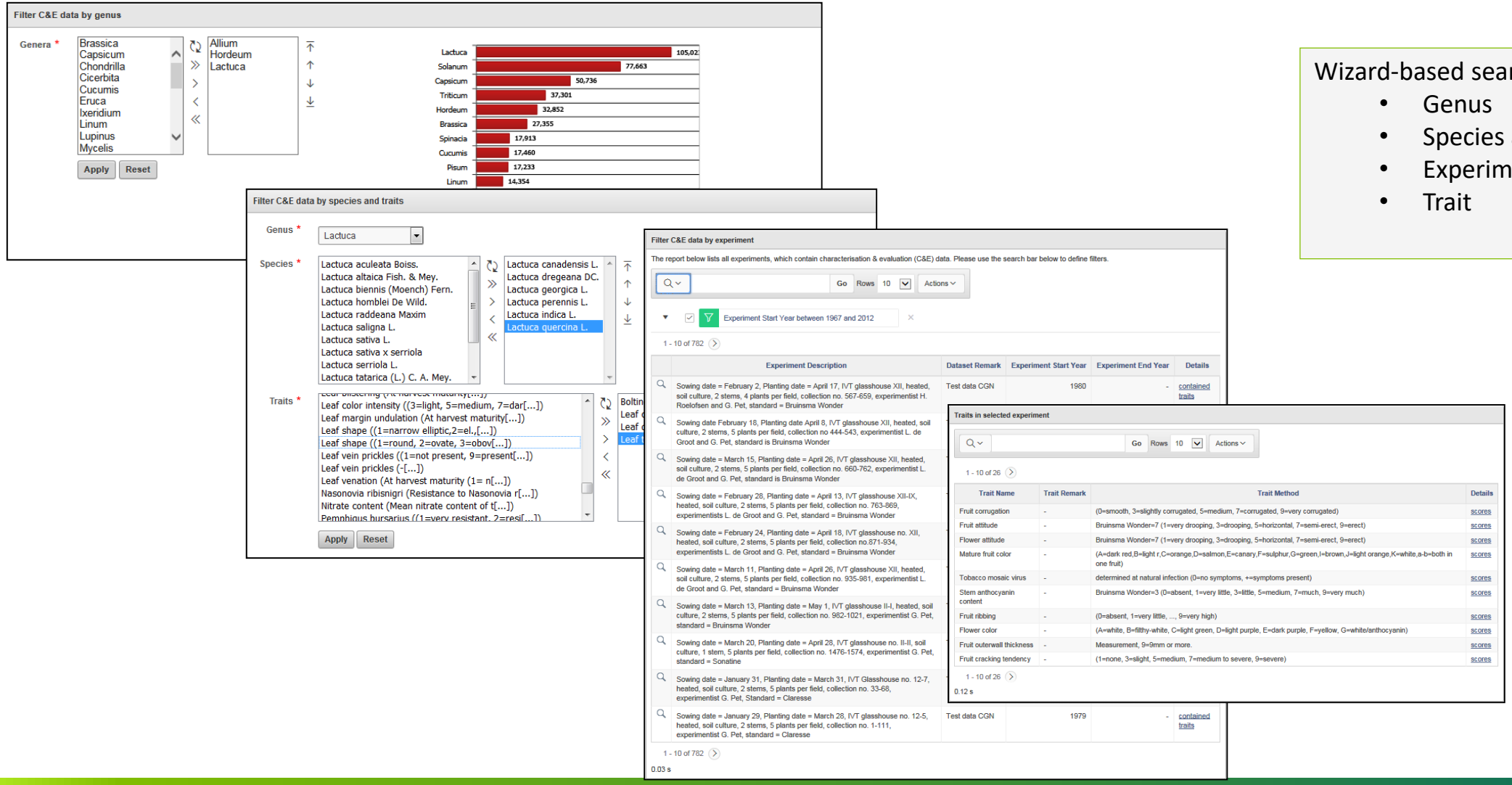
Stephan Weise

Phenotypic data in EURISCO

Status and plans

Third Meeting of the EURISCO Advisory Committee
7th July 2021

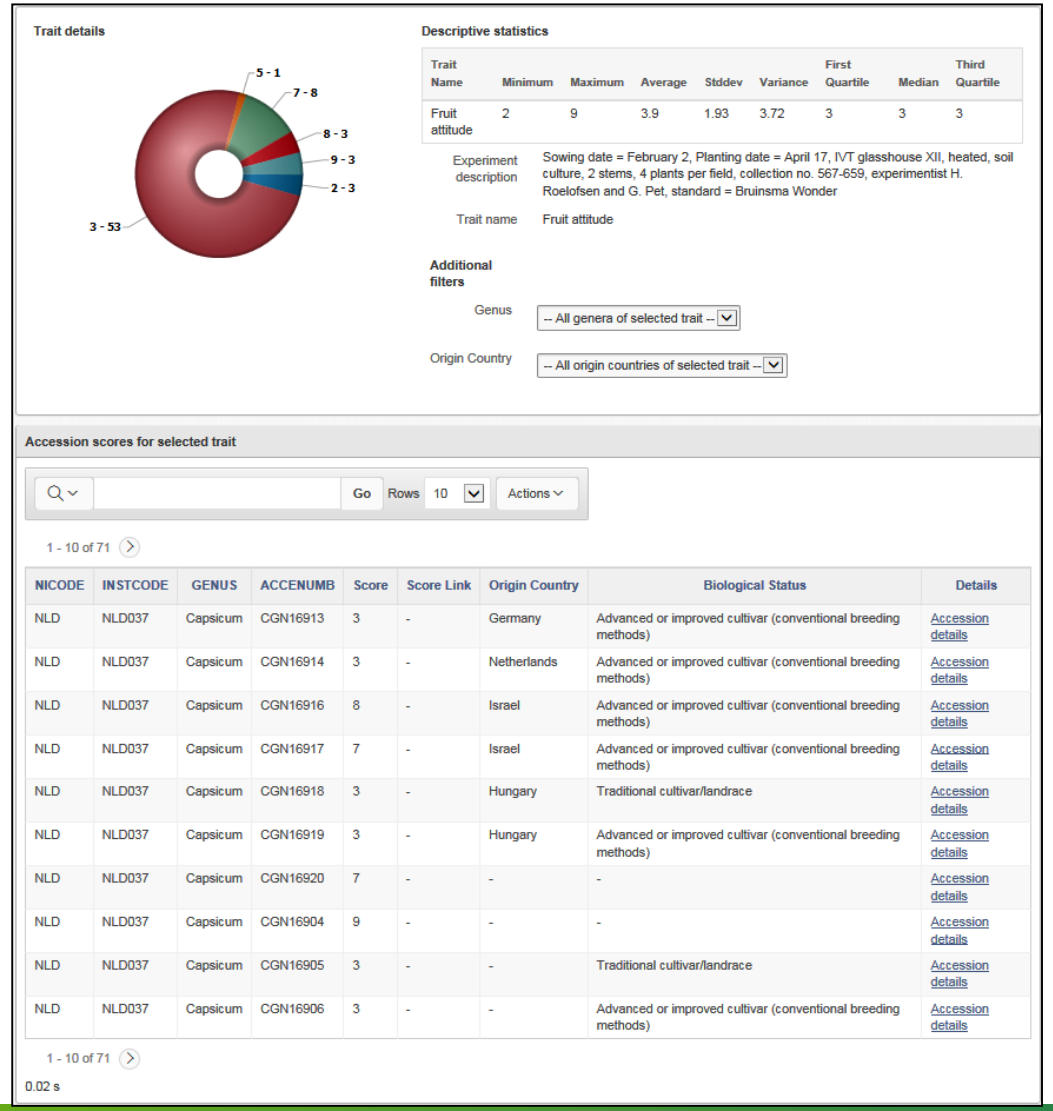
Phenotypic data search in EURISCO



Wizard-based searches for

- Genus
- Species and trait
- Experiment
- Trait

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. Roelofsen and G. Pet, standard = Bruinsma Wonder

Trait name: Fruit attitude

Additional filters

Genus: -- All genera of selected trait --

Origin Country: -- All origin countries of selected trait --

Accession scores for selected trait

Go | Rows 10 | Actions

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

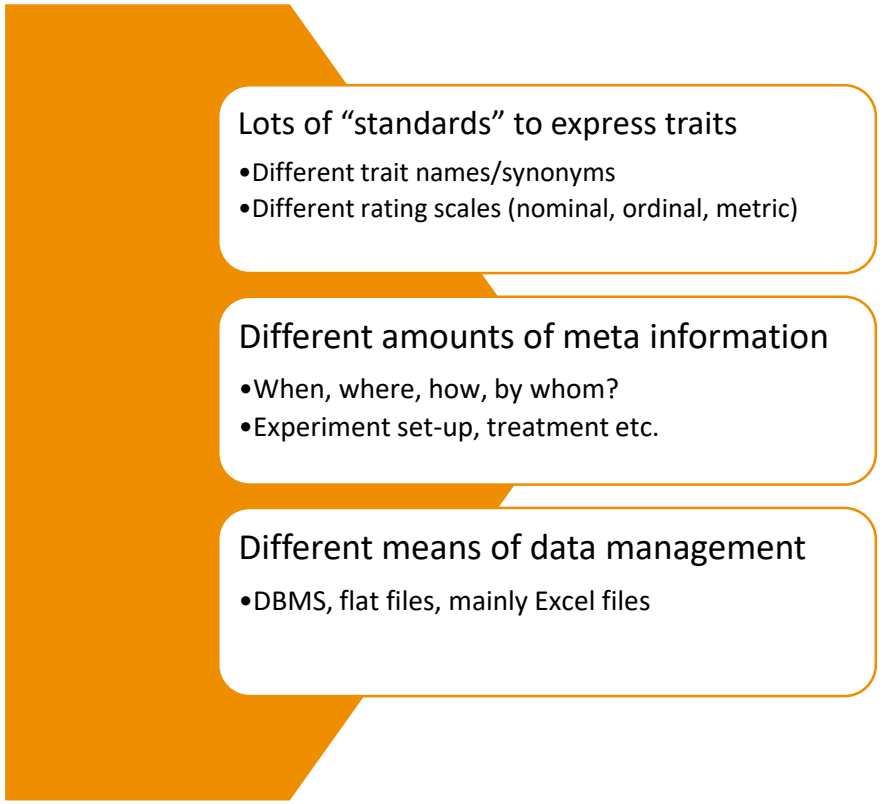
- Genus
- Species and trait
- Experiment
- Trait

Refine result

- Sort
- Filter
- Download
- Chart

Dealing with phenotypic data: Great diversity

- Phenotypic data
  - Determines value of germplasm for breeding and research
  - Crop-specific traits and methods
  - Many historical datasets
  - Usually no data from high throughput phenotyping
  - Data has to be aggregated or exchanged between organisations



Dealing with phenotypic 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)

Dealing with phenotypic data: Current approach

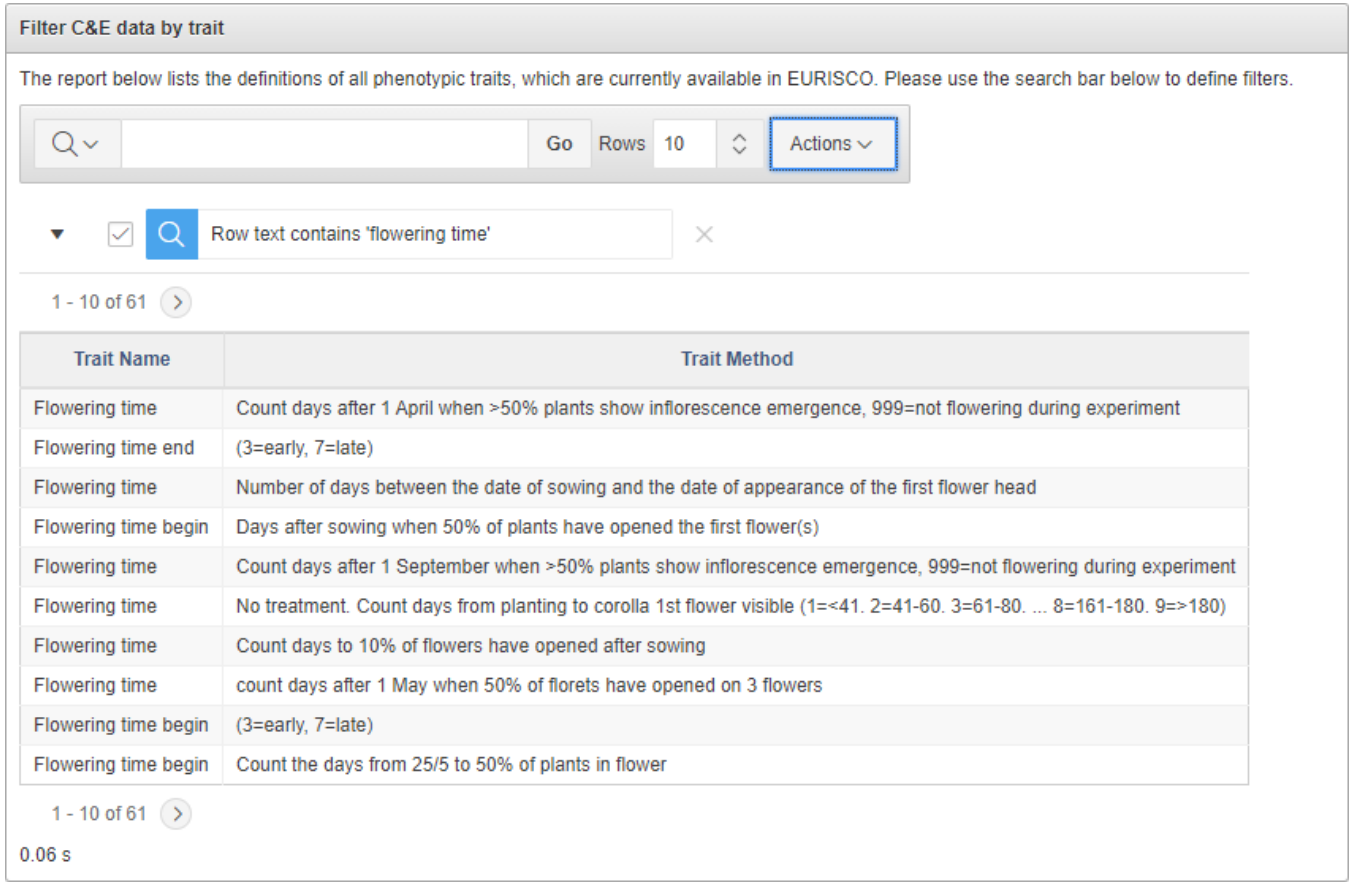
- Data standardisation
  - About 600 germplasm collections in Europe, around 400 in EURISCO
  - No standardisation of trait, scale or experimental design
  - Pragmatic approach: Import of existing data as-is to reach critical mass
- Data exchange
  - Only standardisation of exchange format
    - As simple 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

Dealing with phenotypic data: Data overview

- Extension available since 2016
- 2,680,511 records
- 90,569 accs. with phenotypic data
- 17 countries
- 68 phenotypic datasets
- 3,862 experiments
- 9,415 traits

- Increasingly accepted as repository, but limited comparability



Filter C&E data by trait

The report below lists the definitions of all phenotypic traits, which are currently available in EURISCO. Please use the search bar below to define filters.

Go Rows 10 Actions

Row text contains 'flowering time'

1 - 10 of 61

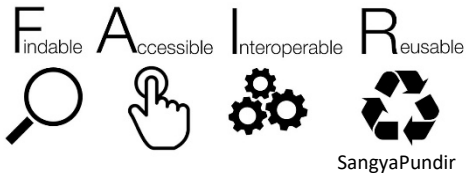
Trait Name	Trait Method
Flowering time	Count days after 1 April when >50% plants show inflorescence emergence, 999=not flowering during experiment
Flowering time end	(3=early, 7=late)
Flowering time	Number of days between the date of sowing and the date of appearance of the first flower head
Flowering time begin	Days after sowing when 50% of plants have opened the first flower(s)
Flowering time	Count days after 1 September when >50% plants show inflorescence emergence, 999=not flowering during experiment
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 to 10% of flowers have opened after sowing
Flowering time	count days after 1 May when 50% of florets have opened on 3 flowers
Flowering time begin	(3=early, 7=late)
Flowering time begin	Count the days from 25/5 to 50% of plants in flower

1 - 10 of 61

0.06 s

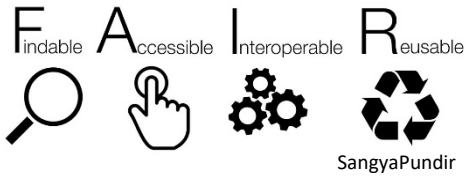
as of 2021-06-23

Dealing with phenotypic data: Towards FAIR data



- Data harmonisation
  - Experiment set-up, treatment etc.
  - Reach MIAPPE-compliance (Krajewski et al. 2015)
- Better structuring
  - Traits/methods/scales
    - Development of common vocabularies/approaches
    - Improve comparability
      - Mapping onto ontology terms
      - Ontology of choice: Crop Ontology (Arnaud et al. 2012)
      - Crux: Sustainability of ontologies
- Provide training + helpdesk
- Additional activities together with various partners, e.g. AGENT

AGENT as a blueprint



- Current limitations
  - EURISCO data exchange format represents a „minimum consensus“
  - Difficult to compile files manually
  - Very limited reproducibility and comparability
- AGENT approach
  - Simplification of data collection → one column per trait to support manual recording
  - Distinction in two types of data

Two “types” of phenotypic data

- Historical phenotypic data
  - Existing data, e.g. from seed multiplication
  - Collected over a long period of time
  - Relatively few metadata available
  - Focus on simple traits

- New experimental data
  - Desirable to describe the experiments as comprehensive as possible
  - Harmonisation of data of the various project partners
    - Traits
    - Methods
    - Experimental set-ups

Two “types” of phenotypic data (cont.)

- Historical phenotypic data
  - Simplified EURISCO-EVA templates should be sufficient

    →Already in use

- New experimental data
  - Templates presently used in EURISCO do not allow to collect additional metadata
    - Only very little information can be represented
    - Focussing on providers managing data in database management systems
  - Strive for increased user-friendliness
  - Consider more extensive metadata
  - Allow full MIAPPE compliance

    → Under development

Separate template for new data?

- Data to be collected
  - Actual observation data → relatively easy to handle
  - Metadata → can be compiled during a datathon
    - Experimental set-up (start and end, plot length and width, number of rows, distance between rows, seeds/m^2, experimental design, …)
    - Treatments (herbicide, fungicide, insecticide, fertiliser, growth regulator, …)
    - Experimental location (location/site name, coordinates, average values of precipitation, temperature, soil type, …)
    - Soil information (N, pH, K_2O, P_2O5, Mg, ….)
    - Traits
      - Definition (name, description, unit, type, method, ontology term)
      - Range of values (rating values, allowed min/max values)

M. Grau / IPK

Thank you for your attention