C&E data templates

Presentation & discussion

EURISCO training workshop, 19th to 21st May 2015, Tirana, Albania
C&E data I

- Currently not available in EURISCO
- Of high importance to users of PGR data
  - Determine value of germplasm for breeding and research
- Difficult to handle due to lots of “standards”
  - Different descriptor names/synonyms
  - Different rating scales
    - Nominal, ordinal, metric scale
  - Different amounts of meta information
    - When, where, how, by whom?
    - Experiment set-up, treatment etc.
C&E data II

- Discussed in detail in previous years
- Suitable proposal compiled by the ECPGR Doc&Info Network in 2009
  - As simple as possible
  - As few fields as possible
    → “minimum consensus”
- D&I meeting 2014
  - Pragmatic approach:
    - No standardisation of trait, scale or experimental design
    - Only standardisation of exchange format (as simple as possible)
    - Import of existing data as-is → reach critical mass
    - Use proposal of 2009 (+ small adjustments)
DATA MODEL
Data model for C&E data I

• Five components
  – GENOTYPE
    • Identified by EURISCO descriptors
  – DATASET
    • May comprise different experiments
  – EXPERIMENT
    • Multiple genotypes are scored for different traits
  – TRAIT
    • Characteristic feature to be scored
  – SCORE
    • Value of a trait for a genotype
TEMPLATES
DATASET

• Enables to upload multiple experiments at once
• Fields:
  – **UPLOADERCODE**: 
    • ID of registered authorised data provider 
    • Provided by EURISCO
  – **DATASET_NUMBER**: 
    • To link experiments with datasets 
    • Unique and persistent for the data provider
  – **DATASET_REMARK**: 
    • General remark for all scores in the dataset

<table>
<thead>
<tr>
<th>UPLOADERCODE</th>
<th>DATASET_NUMBER</th>
<th>DATASET_REMARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>deu_uploader01</td>
<td>1</td>
<td>This dataset contains forage grass accessions.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
EXPERIMENT I

- Meta data helping to interpret C&E data
  - Experiment set-up
  - Weather conditions
  - Soil conditions
  - Experiment location
  - ...

- Fields:
  - **DATASET_NUMBER***:  
    - Reference of the dataset
  - **EXPERIMENT_NUMBER***:  
    - To link scores with experiments
    - Unique and persistent for the data provider
EXPERIMENT II

• Fields (cont.):
  – EXPERIMENT_DESCRIPTION:
    • Brief English description
    • Information necessary for interpreting the scores, e.g. set-up
  – EXPERIMENT_YEAR:
    • Year in which the experiment was performed
  – EXPERIMENT_LONGITUDE:
    • Longitude of experimental site
  – EXPERIMENT_LATITUDE:
    • Latitude of experimental site
  – EXPERIMENT_REPORT:
    • Reference to a report
      – Either report file
      – Or report URL
# Experiment III

<table>
<thead>
<tr>
<th>DATASET_NUMBER</th>
<th>EXPERIMENT_NUMBER</th>
<th>EXPERIMENT_DESCRIPTION</th>
<th>EXPERIMENT_YEAR</th>
<th>EXPERIMENT_LONGITUDE</th>
<th>EXPERIMENT_LATITUDE</th>
<th>EXPERIMENT_REPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Characterisation data of Lolium perenne</td>
<td>1999</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Characterisation data of Lolium perenne</td>
<td>2000</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Characterisation data of Lolium perenne</td>
<td>2001</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Evaluation data of Lolium perenne (4 replications per accession)</td>
<td>2002</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
**Trait I**

- Describe phenotypic traits and the methods used for scoring
- Fields:
  - **Trait_Number***:
    - Unique, temporary number of the trait in the dataset
  - **Trait_Name***:
    - English name of the trait
  - **Trait_Remark***:
    - General remark helping to interpret the trait
  - **Trait_Method***:
    - English description of the used method + scale
## Trait II

<table>
<thead>
<tr>
<th>TRAIT_NUMBER</th>
<th>TRAIT_NAME</th>
<th>TRAIT_REMARK</th>
<th>TRAIT_METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sowing date</td>
<td>...</td>
<td>Date</td>
</tr>
<tr>
<td>2</td>
<td>Emerging date</td>
<td>...</td>
<td>Date</td>
</tr>
<tr>
<td>3</td>
<td>Growing before winter</td>
<td>...</td>
<td>Rating value from 1 (min) – 9 (max)</td>
</tr>
<tr>
<td>4</td>
<td>Stem height min</td>
<td>In flowering time, the shortest plant</td>
<td>Measurement [cm]</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
GENOTYPE I

• All accessions for which C&E data will be uploaded
• Fields:
  – GENOTYPE_NUMBER*: 
    • Unique, temporary number of the genotype in the dataset
  – GENOTYPE_NICODE*: 
    • National Inventory code from EURISCO
  – GENOTYPE_INSTCODE*: 
    • Holding institute code from EURISCO
  – GENOTYPE_ACCENUMB*: 
    • Accession number from EURISCO
  – GENOTYPE_GENUS*: 
    • Genus from EURISCO
## Genotype II

<table>
<thead>
<tr>
<th>Genotype Number</th>
<th>Genotype Nicode</th>
<th>Genotype Instcode</th>
<th>Genotype Accenumb</th>
<th>Genotype Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 142</td>
<td>Lolium</td>
</tr>
<tr>
<td>2</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 476</td>
<td>Lolium</td>
</tr>
<tr>
<td>3</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 550</td>
<td>Lolium</td>
</tr>
<tr>
<td>4</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 2670</td>
<td>Lolium</td>
</tr>
</tbody>
</table>
SCORE I

- Observed phenotypic values of the accessions
- Fields:
  - GENOTYPE_NUMBER*:
    - Reference to a genotype
  - EXPERIMENT_NUMBER*:
    - Reference to an experiment
  - TRAIT_NUMBER*:
    - Reference to a trait
  - SCORE*:
    - Observed score
  - SCORE_LINK:
    - Link to a publication on accession level
## Score II

<table>
<thead>
<tr>
<th>GENOTYPE_NUMBER</th>
<th>EXPERIMENT_NUMBER</th>
<th>TRAIT_NUMBER</th>
<th>SCORE</th>
<th>SCORE_LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>19990313</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>http://...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>20020401</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>21</td>
<td>http://...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
## Putting all this together

### GENOTYPE

<table>
<thead>
<tr>
<th>GENOTYPE_NUMBER</th>
<th>GENOTYPE_NICODE</th>
<th>GENOTYPE_INSTCODE</th>
<th>GENOTYPE_ACCENUMB</th>
<th>GENOTYPE_GENUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 142</td>
<td>Lolium</td>
</tr>
<tr>
<td>2</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 476</td>
<td>Lolium</td>
</tr>
<tr>
<td>3</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 550</td>
<td>Lolium</td>
</tr>
<tr>
<td>4</td>
<td>DEU</td>
<td>DEU271</td>
<td>GR 2670</td>
<td>Lolium</td>
</tr>
</tbody>
</table>

### TRAIT

<table>
<thead>
<tr>
<th>TRAIT_NUMBER</th>
<th>TRAIT_NAME</th>
<th>TRAIT_REMARK</th>
<th>TRAIT_METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sowing date</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>Emerging date</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>Growing before winter</td>
<td>...</td>
<td>Rating value from 1 (min) – 9 (max)</td>
</tr>
<tr>
<td>4</td>
<td>Stem height min</td>
<td>In flowering time, the shortest plant</td>
<td>Measurement [cm]</td>
</tr>
</tbody>
</table>

### SCORE

<table>
<thead>
<tr>
<th>GENOTYPE_NUMBER</th>
<th>EXPERIMENT_NUMBER</th>
<th>TRAIT_NUMBER</th>
<th>SCORE</th>
<th>SCORE_LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>19990313</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>http://...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>20020401</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>21</td>
<td>http://...</td>
</tr>
</tbody>
</table>

### EXPERIMENT

<table>
<thead>
<tr>
<th>DATASET_NUMBER</th>
<th>EXPERIMENT_NUMBER</th>
<th>EXPERIMENT_DESCRIPTION</th>
<th>EXPERIMENT_YEAR</th>
<th>EXPERIMENT_LONGITUDE</th>
<th>EXPERIMENT_LATITUDE</th>
<th>EXPERIMENT_REPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Characterisation data of Lolium perenne</td>
<td>1999</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Characterisation data of Lolium perenne</td>
<td>2000</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Characterisation data of Lolium perenne</td>
<td>2001</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Evaluation data of Lolium perenne (4 replications per accession)</td>
<td>2002</td>
<td>11.278414</td>
<td>51.826059</td>
<td>http://...</td>
</tr>
</tbody>
</table>

### DATASET

<table>
<thead>
<tr>
<th>UPLOADERCODE</th>
<th>DATASET_NUMBER</th>
<th>DATASET_REMARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>deu_uploader01</td>
<td>1</td>
<td>This dataset contains forage grass accessions.</td>
</tr>
</tbody>
</table>

---

**EURISCO**

Finding seeds for the future

**LEIBNIZ INSTITUTE OF PLANT GENETICS AND CROP PLANT RESEARCH**

**IPK**

Gatersleben
Proceeding for data upload

• **Prerequisite:**
  – Only non-confidential C&E data
  – Only data of accessions listed in EURISCO

• **Impact**
  – NFPs responsible for data upload
    → May nominate users for (sub) accounts
  – Descriptor HISTORIC
    → Currently under discussion

• **Upload via Java tool**
  → Coming soon