C&E data templates

Presentation & discussion

EURISCO training workshop, 19th to 21st May 2015, Tirana, Albania

Stephan Weise
21 May 2015

BACKGROUND

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.
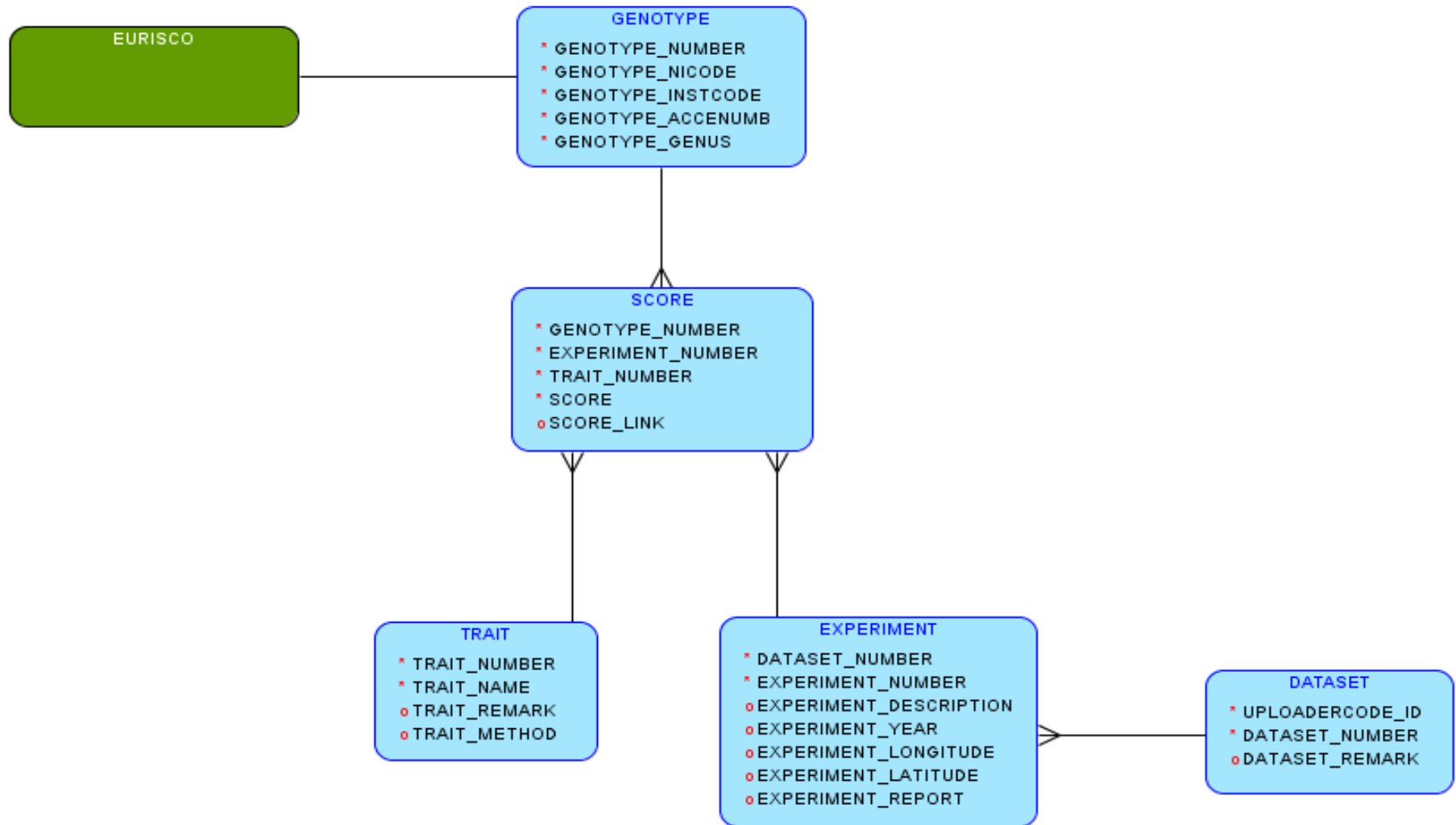
C&E data II

- 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

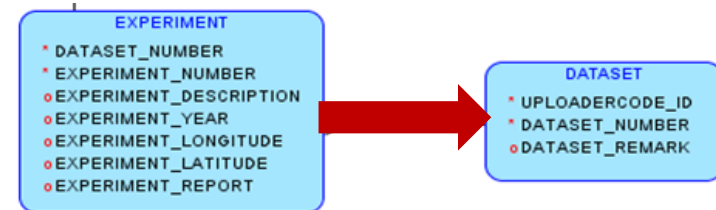# Data model for C&E data II

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

UPLOADERCODE	DATASET_NUMBER	DATASET_REMARK
deu_uploader01	1	This dataset contains forage grass accessions.
…	…	…
…	…	…

eurisco
Finding seeds for the future

IPK
GATERSLEBEN

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
* DATASET_NUMBER
* EXPERIMENT_NUMBER
o EXPERIMENT_DESCRIPTION
o EXPERIMENT_YEAR
o EXPERIMENT_LONGITUDE
o EXPERIMENT_LATITUDE
o EXPERIMENT_REPORT

DATASET
* UPLOADERCODE_ID
* DATASET_NUMBER
o DATASET_REMARK

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

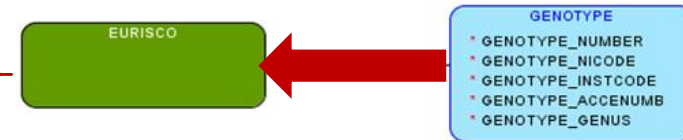| DATASET_NUMBER	EXPERIMENT_NUMBER	EXPERIMENT_DESCRIPTION	EXPERIMENT_YEAR	EXPERIMENT_LONGITUDE	EXPERIMENT_LATITUDE	EXPERIMENT_REPORT
1	1	Characterisation data of Lolium perenne	1999	11.278414	51.826059	http://...
1	2	Characterisation data of Lolium perenne	2000	11.278414	51.826059	http://...
1	3	Characterisation data of Lolium perenne	2001	11.278414	51.826059	http://...
1	4	Evaluation data of Lolium perenne (4 replications per accession)	2002	11.278414	51.826059	http://...
...	...	...	...	...	...	...

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

TRAIT_NUMBER	TRAIT_NAME	TRAIT_REMARK	TRAIT_METHOD
1	Sowing date	…	Date
2	Emerging date	…	Date
3	Growing before winter	…	Rating value from 1 (min) – 9 (max)
4	Stem height min	In flowering time, the shortest plant	Measurement [cm]
…	…	…	…

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

EURISCO

GENOTYPE
* GENOTYPE_NUMBER
* GENOTYPE_NICODE
* GENOTYPE_INSTCODE
* GENOTYPE_ACCENUMB
* GENOTYPE_GENUS

GENOTYPE II

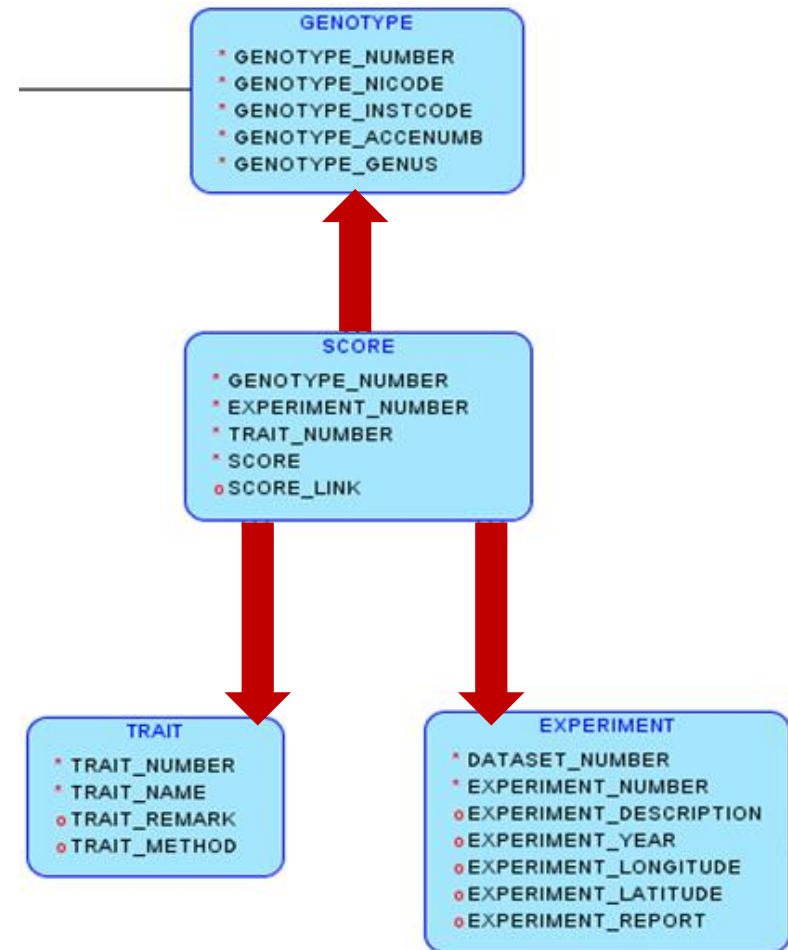| GENOTYPE_NUMBER	GENOTYPE_NICODE	GENOTYPE_INSTCODE	GENOTYPE_ACCENUMB	GENOTYPE_GENUS
1	DEU	DEU271	GR 142	Lolium
2	DEU	DEU271	GR 476	Lolium
3	DEU	DEU271	GR 550	Lolium
4	DEU	DEU271	GR 2670	Lolium

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

GENOTYPE
- * GENOTYPE_NUMBER
- * GENOTYPE_NICODE
- * GENOTYPE_INSTCODE
- * GENOTYPE_ACCENUMB
- * GENOTYPE_GENUS

SCORE
- * GENOTYPE_NUMBER
- * EXPERIMENT_NUMBER
- * TRAIT_NUMBER
- * SCORE
- o SCORE_LINK

TRAIT
- * TRAIT_NUMBER
- * TRAIT_NAME
- o TRAIT_REMARK
- o TRAIT_METHOD

EXPERIMENT
- * DATASET_NUMBER
- * EXPERIMENT_NUMBER
- o EXPERIMENT_DESCRIPTION
- o EXPERIMENT_YEAR
- o EXPERIMENT_LONGITUDE
- o EXPERIMENT_LATITUDE
- o EXPERIMENT_REPORT

Score II

GENOTYPE_NUMBER	EXPERIMENT_NUMBER	TRAIT_NUMBER	SCORE	SCORE_LINK
1	1	1	19990313	http://...
1	1	3	7	http://...
4	4	1	20020401	...
4	4	4	21	http://...
...	...	...	...	...
...	...	...	...	...

Putting all this together

GENOTYPE

GENOTYPE_NUMBER	GENOTYPE_NICODE	GENOTYPE_INSTCODE	GENOTYPE_ACCENUMB	GENOTYPE_GENUS
1	DEU	DEU271	GR 142	Lolium
2	DEU	DEU271	GR 476	Lolium
3	DEU	DEU271	GR 550	Lolium
4	DEU	DEU271	GR 2670	Lolium

TRAIT

TRAIT_NUMBER	TRAIT_NAME	TRAIT_REMARK	TRAIT_METHOD
1	Sowing date	…	Date
2	Emerging date	…	Date
3	Growing before winter	…	Rating value from 1 (min) – 9 (max)
4	Stem height min	In flowering time, the shortest plant	Measurement [cm]
…	…	…	…

SCORE

GENOTYPE_NUMBER	EXPERIMENT_NUMBER	TRAIT_NUMBER	SCORE	SCORE_LINK
1	1	1	19990313	http://…
1	1	3	7	http://…
4	4	1	20020401	…
4	4	4	21	http://…
…	…	…	…	…
…	…	…	…	…

EXPERIMENT

DATASET_NUMBER	EXPERIMENT_NUMBER	EXPERIMENT_DESCRIPTION	EXPERIMENT_YEAR	EXPERIMENT_LONGITUDE	EXPERIMENT_LATITUDE	EXPERIMENT_REPORT
1	1	Characterisation data of Lolium perenne	1999	11.278414	51.826059	http://…
1	2	Characterisation data of Lolium perenne	2000	11.278414	51.826059	http://…
1	3	Characterisation data of Lolium perenne	2001	11.278414	51.826059	http://…
1	4	Evaluation data of Lolium perenne (4 replications per accession)	2002	11.278414	51.826059	http://…
…	…	…	…	…	…	…

DATASET

UPLOADERCODE	DATASET_NUMBER	DATASET_REMARK
deu_uploader01	1	This dataset contains forage grass accessions.
…	…	…
…	…	…

PROCEEDING

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