Phenotype data: Challenges for AgBioData databases

AgBioData Webinar 8/2/2023

Today's agenda

- Introduction to the problem (10')
- Breakout rooms (30')
- Report back from the breakout room (2-3' per room - total 10')
- Conclusions/next steps (10')

Today's Goals

1. Describe challenges for AgBioData databases regarding phenotypic data;
2. Identify the most pressing challenges;
3. Identify whether any of these challenges could be addressed by a working group in the coming year.

Phenotypic data challenges for AgBioData databases

- Curation, storage, and access/sharing of phenotypic data presents challenges for multiple AgBioData databases.
- In last year's Data Federation survey, seven out of 38 databases stated that phenotypic and phenomic data are still challenging to share;
- Last year's call for working group proposals included five proposals that addressed phenotypic/phenomic data at least in part.
- There is clearly a need to address challenges with phenotypic data, but the challenges are diverse.
- Given the complexity of challenges around phenotypic data for AgBioData databases, can we as a group identify a topic that a working group could address?

Informal assessment results

- We asked 14 AgBioData databases what phenotypic data types they curated, stored or accessed, and what challenges they experienced with phenotypic data.
- We learned the following main concerns from the 8 responses:
  - ***Data type diversity***: images, tables with measurements, etc.
  - ***Data source diversity (where the DBs get the data from):*** breeders, researchers, literature
  - ***Data Standardization:*** it is difficult to integrate trait observations across species and field trials because of the lack of ontologies or controlled vocabularies that can be used at least in some steps of the phenotypic process.
  - ***Lack of a centralized database*** to share plant image data

Discussion questions

- Have each participant briefly (under 2 minutes) answer the following:
    - What database do you work for or use?
    - Does your database store/curate/access phenotypic data, and if so, which?
    - If yes, please describe 1-3 challenges your database faces regarding the curation, storage, and/or access of phenotypic data.
- How relevant are the following issues for your database in terms of phenotypic data?
    - lack of standardization/ontologies across and within species makes phenotypic data curation hard and time-consuming.
    - data type diversity makes curation, storage and/or access challenging (e.g, spreadsheet, images, etc.)
    - data source diversity (where the DBs get the data from; e.g., breeders, literature, researchers, etc.)
    - lack of a centralized database
- Based on your discussion, what challenges do you think we can address completely or partially within a working group? Are there other efforts we should join or involve in this project?
- Would you like to join a phenotype working group if established? Would you nominate yourself or someone else as WG chair or champion?