FAIR from the Start: Empowering Data Sharing in Agricultural Research

FAIR Scientific Literature Working Group Update

AgBioData Workshop

PAG33

1/11/2026

Leonore Reiser

Members

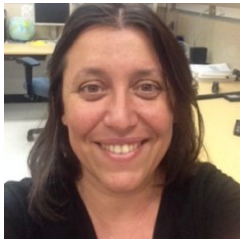
Leyla Cabugos
Librarian, Cal Poly

Jacqueline Campbell
Geneticist, USDA ARS
SoyBase curator

Jenna Daenzer
Peer review Coordinator
Genetics Society of America

Sook Jung
Asst Research Professor. GDR

David Molik
Center for Scholarly Publishing, Kansas State University
Department head



Daniela Raciti
Exec Editor, microPublication & Wormbase Curator

Leonore Reiser
TAIR curator

Adam Wright
Software Engineer
WormBase, Reactome

Karen Yook
Exec Editor, microPublication & Wormbase Curator

We gratefully acknowledge support from NSF, grant #2126334

FSL Working Group Goals

✓ Identify bottlenecks in the publication-curation pipeline.

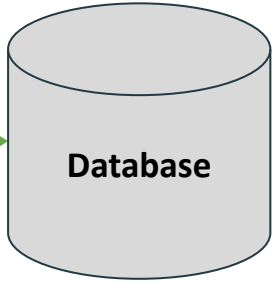
- Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.
- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.

A better literature curation workflow

AUTHOR
Data generator

Database
CURATOR

- ✓ Data curation at time of publication
- ✓ Clear guidance on data submission & databases
- ✓ Validators to check for data submission
- ✓ Validators for data formats
- ✓ Nomenclature standards and validators
- ✓ No more data available upon request
- ✓ Better paper metadata, AI based paper classification
- ✓ More funding for curation, AI curation assistants



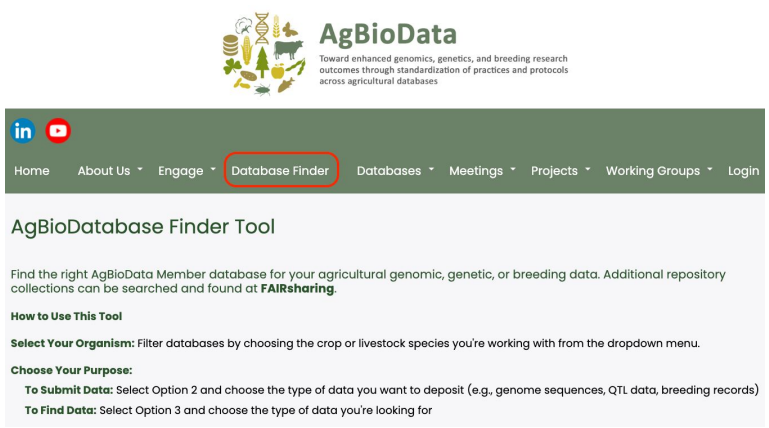
FSL Working Group Goals

- Identify bottlenecks in the publication-curation pipeline.

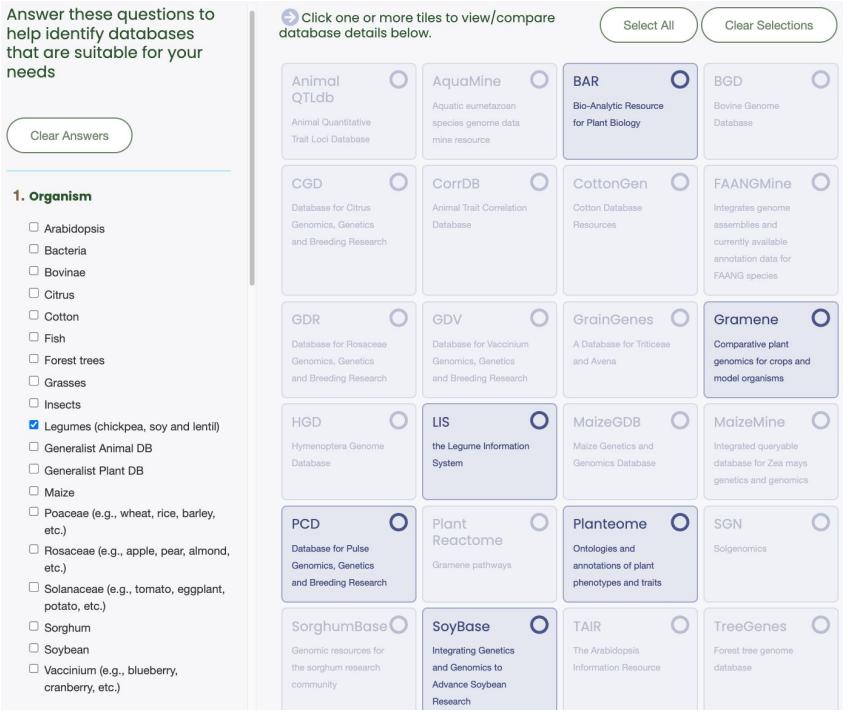
- ✓ Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.

- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.

Tool: AgBio Database Finder Drupal Module



Option 1:
Filter by species



https://www.agbiodata.org/databasefinder
https://doi.org/10.17912/micropub.biology.001896

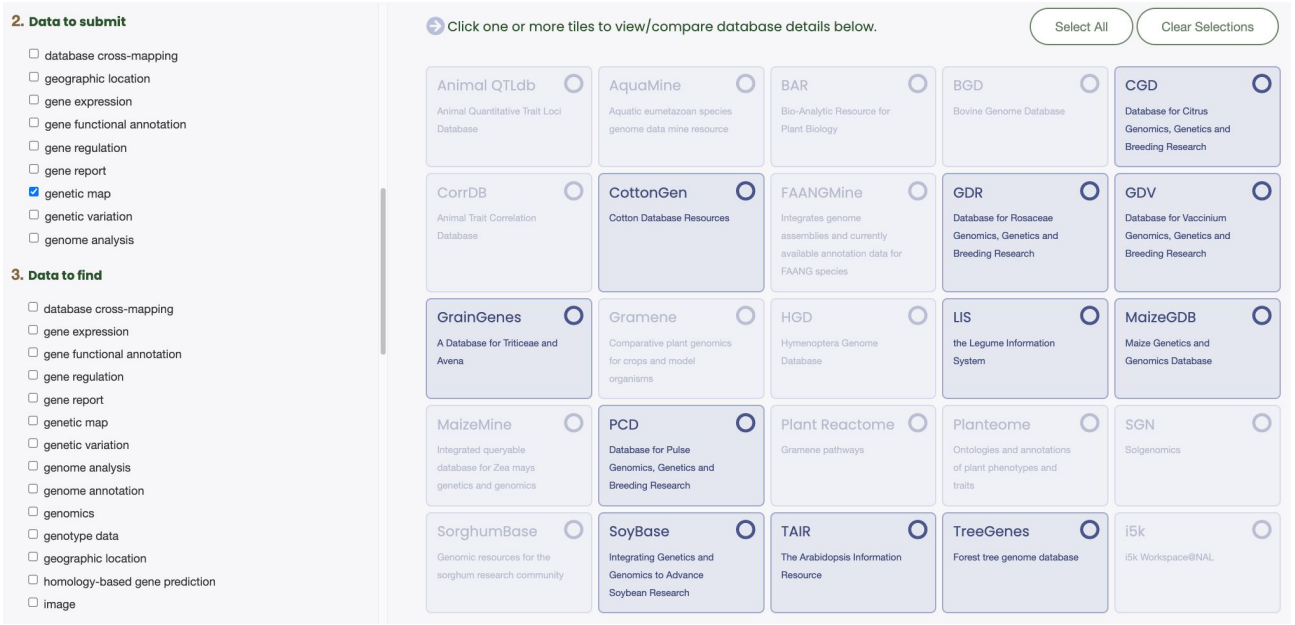
https://github.com/CU-CommunityApps/CD-finder

Option 2: Filter by …

Data to submit

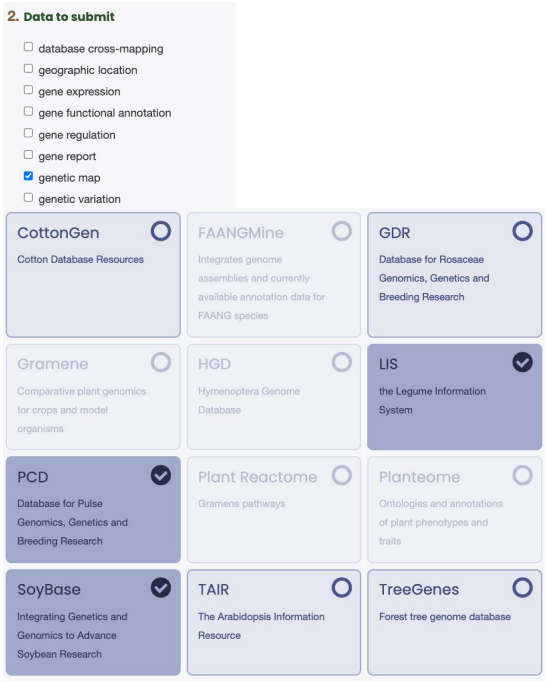
OR

Data to find



https://www.agbiodata.org/databasefinder

Apply filters and compare resources



LIS PCD SoyBase

	LIS	PCD	SOYBASE
DB name	LIS - the Legume Information System	Pulse Crop Database (PCD)	SoyBase
DB short description	LegumeInfo is a federally funded informatics service assisting in basic research in the legumes by relating data from multiple species, and traverse among various data types.	The Pulse Crop Database (PCD), formerly the Cool Season Food Legume Database (CSFL), has been developed by the Main Bioinformatics Laboratory at Washington State University in collaboration with the USDA-ARS Grain Legume Genetics and Physiology Research Unit, the USDA-ARS Plant Germplasm Introduction and Testing Unit, the USA Dry Pea and Lentil Council, Northern Pulse Growers and allied scientists in the US and across the world, to serve as a resource for Genomics-Assisted Breeding (GAB). GAB offers tools to identify genes related to traits of interest, among other methods to optimize plant breeding efficiency and research, by providing relevant genomic, genetic, and breeding information and analysis. Therefore, tools such as JBrowse and MapViewer can be found in this database, as well as key resources to provide access to the annotation of available transcriptome data, helping pulse breeders and researchers to succeed in their programs.	SoyBase is a long-term, federally funded genetics and breeding database for soybean breeders and researchers focused on soybeans.
DB URL	https://www.legumeinfo.org/	https://www.pulsedb.org/	https://www.soybase.org/
DB data types	• database cross-mapping* • gene expression • gene functional annotation* • gene regulation • gene report • genetic map* • genetic variation • genome analysis • genome annotation* • genomics • genotype data* • geographic location • homology-based gene prediction* • metadata • molecular sequence annotation* • phylogenetic data* • QTL map* • sequence features* • whole genome association study*	• database cross-mapping • gene expression* • gene functional annotation* • gene regulation* • gene report* • genetic map* • genetic variation* • genome analysis • genome annotation* • genotype data* • homology-based gene prediction* • image* • metabolic pathway prediction • ontology and terminology • phenotype* • QTL* • whole genome association study* * = Direct submission accepted	• database cross-mapping* • gene expression • gene functional annotation* • genetic map* • gene regulation • gene report • genetic variation • genome analysis • genome annotation • genomics • genotype data • geographic location • homology-based gene prediction • metabolic pathway prediction • metadata • nucleic acid sites, features, and motifs • ontology and terminology • pedigree* • phenotype*

https://www.agbiodata.org/databasefinder

What about source/data type repositories?

- Add more tiles/databases to the tool?
- Create a collection in FAIRsharing?
- Maintain a separate list ordered by data type?
- *AI/Chatbot?*

(How) can AI (tools) help make published data more FAIR?

- What is the future of publishing and co-ordinated data curation in the age of AI?
- Can we make the data and metadata in papers more machine readable/FAIR?
- A freely accessible, universal AI tool that:
  - Scans papers and identifies types of data
  - Extracts experiment metadata
  - Validates data and metadata formats
  - Checks against nomenclature and other standards
  - Readies data for submission to appropriate repository

How to engage with publishers and journals?

How to introduce data curation into the publishing pipeline

How to connect journals to community databases

Incentives

Apicuron Credits for authors

FAIRness badges for papers/journals

Workflow changes

Pre-submission or pre-print curatorial review

Advanced vetting to speed peer review

Start with society journals

Journals need assurance that recommended repos are FAIR

Increasing FAIRness of AgBioData Member Databases

- FAIR sharing record completion
- Assess FAIRness of databases
  - What minimum expectation and standards for member databases
  - How can AgBioData support meeting those expectation?

FSL Working Group Goals

- Identify bottlenecks in the publication-curation pipeline.
- Identify sets of existing or desired tools or biocuration resources to increase literature curation throughput and accuracy.
- Publish recommendations and a roadmap for authors and publishers to increase the FAIRness of research.

Thoughts for Discussion

- What if the roadmap was an *interactive tool* not a static white paper?

- How could an integrated curation- publication data pipeline be made sustainable?