AgBioData Consortium

2024 Survey of Genomic, Genetic, and Breeding (GGB) Database Team Members

Summary of 2024 Follow-up Data

v1.4 (9/19/2024)

Survey Sample, Participant Characteristics, and Familiarity and Experience with GGB Databases

A total of 38 usable survey responses were received during Spring 2024, with complete or mostly complete responses to the substantive questions that were asked after the initial self-descriptors at the beginning of the survey. Sample sizes for specific questions may vary slightly since a few people left particular questions unanswered. In some of the following tables, reported percentages may not sum to 100 due to rounding.

The responding group of database team members was not a random sample of all team members, since the invitation to participate was circulated widely, and those who responded were self-selected, i.e. they chose to participate. We can't know the extent to which this group may be representative of all team members, so we can't confidently generalize the responses of this group to the entire population of GGB database team members. We can, however, be confident that we know with reasonable clarity what this group of 38 GGB database team members reported in 2024 about their experiences, opinions, and recommendations concerning these databases.

Direct comparisons with the 2022 baseline survey must be interpreted cautiously because of both sampling and measurement issues. The 2022 sample was also an anonymous, non-random group that, to an unknown extent, may overlap with the set of individuals who chose to participate in 2024. At the request of the project team, response options and/or wording of some questions were altered for the 2024 follow-up survey; findings for these questions are thus not exactly comparable with the 2022 baseline survey.

Over a quarter of respondents (26%) reported working in a university that offers related PhD degrees; 24% reported working in a Land Grant university; 45% reported working for the US Department of Agriculture (see Table 1). Others reported working for other institutions: another government agency, an intergovernmental organization, or a public research organization. As displayed in Table 2, most reported their primary professional role as being a research scientist (32%), faculty member (21%), biocurator (18%), software developer (13%), bioinformatician/data analyst (10%), or postdoc/graduate student (6%). (The response options for this question differed slightly in 2024 compared to the baseline 2022 survey.)

Most respondents reported a professional focus on plants, including major plant crops (63%), model organisms (37%), horticultural specialty crops (24%), plants grown for purposes besides human consumption (34%), and wild organisms not directly used in agriculture (24%); see Table 3. Animals were a primary focus for fewer respondents, including major livestock animals (13% of respondents), minor livestock animals (16%), model organisms (10%), and wild organisms not directly used in agriculture (5%). Pests, diseases, physiological stressors, and other threats were a focus for 16% of respondents. "Other" areas of professional focus were insects, data standards, ontologies, and software/platform development. Respondents could indicate more than one area of focus. (The response options for this question differed slightly in 2024 compared to the baseline 2022 survey.)

Table 1. Organizational Affiliations of 2024 Survey Respondents

Which terms best describe your organization? *(mark all that apply)*	
45%	US Department of Agriculture (USDA)
3%	Other government agency or department
24%	Land Grant University
26%	University offering related PhD degrees
0%	University offering related Masters (but not PhD) degrees
0%	Primarily undergraduate institution (PUI)
0%	Minority-serving institution (MSI)
0%	Historically Black College or University (HBCU)
0%	Private company or industry organization
24%	Nonprofit organization
10%	Other organization *("From university our the United States" "intergovernmental organization" "Public research organization")*

Table 2. Primary Professional Role of 2024 Survey Respondents

Please indicate your primary role:	
32%	Research scientist
21%	Faculty (e.g. Assistant Professor, Adjunct Professor, Professor, etc.)
18%	Biocurator
13%	Software developer
10%	Bioinformatician / Data Analyst
0%	Technician
0%	Librarian
3%	Post-doc
3%	Graduate student
0%	Undergraduate student
0%	Other

Table 3. Professional Focus of 2024 Survey Respondents

Please indicate your primary area(s) of professional focus:	
63%	Plants: Major plant crops grown as food for people on a large percentage of farm land (e.g. maize, soybeans, rice etc.)
24%	Plants: Horticultural, specialty, or other plant crops grown as food for people on a smaller percentage of farm land (e.g. vegetables, fruits, nuts, etc.)
37%	Plants: Model organisms
34%	Plants grown for animal feed, fiber, lumber, industrial use, or ecosystem services/cover crops
24%	Plants: Wild organisms not directly used in agriculture (as potential germplasm sources for agriculture or for understanding biology and ecosystems independent of potential agricultural applications).
13%	Animals: Major livestock animals grown as food for people on a large percentage of ranch/farm land (e.g. beef or dairy cattle, pork, chicken, turkey)
16%	Animals: Minor livestock animals such as sheep or goats; fish, shellfish or other aquatic animals raised as food for people; honey bees, etc.
10%	Animals: Model organisms
5%	Animals: Wild organisms not directly used in agriculture (as potential germplasm sources for agriculture or for understanding biology and ecosystems independent of potential agricultural applications)
16%	Pests, diseases, physiological stressors, other threats to agriculture or ecosystems
10%	Other: *"Animal: Insects" "Data Standards" "Ontologies" "Software/platform development"*

As displayed in Table 4, 95% of participants "strongly" or "moderately" agreed that they are "very familiar with the concept of FAIR data and could explain them to others," while 79% "strongly" or "moderately" agreed that they are "very familiar with technical/practical tools for FAIR data management and use them in my work." 82% "strongly" or "moderately" agreed that "AgBioData has helped me to become more familiar and experienced with FAIR data management principles and resources"; 42% "strongly" or "moderately" agreed that "I rely on other resources to learn about FAIR data management principles and resources." (The wording of the first question remained identical to that used at baseline, but the other questions in this section were newly phrased for the 2024 survey.)

Table 4. Team Member Familiarity with Implementation of FAIR Data Practices, 2024

Based on your knowledge of and experience with FAIR data and GGB databases, please rate how much you agree with these statements:	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
I am very familiar with the concept of FAIR data principles and could explain them to others.	-	3 %	-	3	29	66
I am very familiar with technical/practical tools for FAIR data management and use them in my work.	-	5	-	16	42	37
AgBioData has helped me to become more familiar and experienced with FAIR data management principles and resources.	-	3	3	13	37	45
I rely on other resources to learn about FAIR data management principles and resources.	-	8	18	32	26	16

Please add comments, extensions, clarifications or recommendations related to any of the questions above:

- *I have learned a lot about FAIR and what it means and doesn't. I think most if not all scientists agree with FAIR principles but in fact need more funding and sustained database support to move forward. We need to make federally-funded databases a resource that is available, free to users and comprehensive.*
- *I interpreted the question "I rely on other resources to learn about FAIR data management principles and resources" as "I also rely on other resources to learn about FAIR data management principles and resources."*
- *Got to know about FAIR data while working as a member of AgBioData working groups.*
- *I was one of the first to talk about FAIR data about 8 years ago in AgBioData meetings ;)*
- *AgBioData has really helped with education on FAIR and other data management related issues.*

Note. N = 38. Each row of numbers contains the proportion (percentage) of participants who gave each response to the question. Row percentages may not add to exactly 100% due to rounding.

2024 Ratings and Recommendations On FAIR Data Practices in GGB Databases and Resources

GGB database or resource team members were asked to rate their agreement with statements about their experiences, observations and opinions regarding the current status of FAIR data practices in GGB databases. They were also asked to rate priorities for future development of FAIR data practices in these projects and to provide related comments and recommendations. The distributions of their responses are displayed in Tables 5 – 7. Note that some of these questions were also asked in the 2022 baseline survey using identical or near-identical wording; not all questions from 2022 were repeated in 2024. The discussion below highlights differences from 2022 to 2024. Although in most cases the 2024 ratings are more favorable with regard to FAIR data practices than the 2022 ratings, keep in mind that this could be due to sampling error (different individuals from different database and resource projects choosing to participate in 2024 vs. 2022) rather than measurable change or improvement across this time period.

Questions C1, C2 and C3 in Table 5 ask how well projects are maintaining consistent data practices. 2024 ratings for standard nomenclatures, metadata, and common file formats ranged from 79% to 92% of respondents "strongly" or "moderately" agreeing that their projects apply these practices consistently. In 2022, baseline survey participants gave lower ratings to their projects (59% to 76%) for these questions.

Questions C4 and C5 ask about user-friendliness of data system tools for contributing and finding/retrieving data. In 2024, ratings for ease of contributing FAIR and accurate data and ease of finding and retrieving data showed 58% and 66% of respondents "strongly" or "moderately" agreeing that their projects provide this for users. In 2022, baseline survey participants gave comparable ratings to their projects for data contribution (59%) and lower ratings for ease of data retrieval (55%).

Questions C6 and C7 in Table 5 ask about 2024 perceptions of the extent to which the GGB databases provide good guidelines for users on some key FAIR data practices. Regarding metadata and data process transparency, 68% of respondents "strongly" or "moderately" agreed that "The project(s) I work for provide good guidelines on what metadata to provide when preparing/submitting data" and 59% of respondents "strongly" or "moderately" agreed that "The project(s) I work for provide good guidelines on how to provide information about how the data have been generated and cleaned, how missing data has been handled, and other process transparency issues." In 2022, baseline survey participants gave lower ratings to their projects (54% and 37% respectively) for these two questions.

Questions C8 and C9 pertain to education of data system users. In 2024, 60% of respondents "strongly" or "moderately" agreed that "the project(s) I work for provide useful resources for learning how to use them: tutorials, FAQs, how-to documents and videos, etc." The corresponding ratings from the 2022 survey sample were 76%, indicating that the 2024 sample was less likely to report that their project provided helpful educational resources for users. In 2024, 53% of respondents "strongly" or "moderately" agreed that "the project(s) I work for highlight the importance of FAIR data principles for researchers and provide educational resources to help researchers understand and follow FAIR practices." The corresponding figure from the 2022 survey sample was 44%. Projects represented in the 2024 sample appear more likely to educate users about FAIR practices, but less likely to provide helpful education about how to use their own tools and resources, compared to the projects sampled in 2022.

On question C10, 87% of 2024 respondents "strongly" or "moderately" agreed that their project employs "cost-sharing efficiencies such as reusable open-source software." Similar ratings (84%) were reported in 2022. Finally, question C11 asked respondents to rate the extent to which "reliable interconnectedness and interoperability between related databases make it easy to combine and integrate data in new ways to address new questions." In 2024 34% of respondents "strongly" or "moderately" agreed, compared to 20% in 2022.

Response patterns to the questions in Table 5 may provide insight into the relative strengths and weaknesses of GGB data projects as perceived by team members. As in 2022, the 2024 sample gave relatively high ratings to the use of cost-sharing efficiencies such as open source software the consistent use of standard nomenclatures, and the use of common file formats. Relatively low ratings were given for documentation of how the data have been generated and cleaned, how missing data has been handled, and other process transparency issues, along with interconnectedness and interoperability between related databases.

There are various ways to scan these ratings to identify which areas appear to be most ripe for improvement. We could add the two (or three) right-most columns to produce a quick sum that would show those items with the highest ratings (GGB database strengths), or conversely, add the two (or three) left-most columns to produce a quick sum that would show those items with the lowest ratings (weaknesses). For now, a quick color-coding scheme has been applied to give some indication of how this might be interpreted; darker blue indicates relative strengths, darker red indicates relative weaknesses that could be important targets for improvement.

Participants were offered a chance to make open-ended comments on the issues; four comments were received:

- *Most databases are not fully supported, and we need more coordinated funded federal resources.*
- *I was able to understand that data federation is important for efficient FAIR policy.*
- *There is no "universal" or recommended way to exchange data short of the BrAPI convention.*
- *Interoperability of data and consistency in how data is submitted are still challenges for our database. The biggest help was journals requiring data submission. This allowed our team to set metadata standards and to convince more users to upload/ share their own data to reduce our curator's workload.*

Table 5. 2024 Appraisal of the Status of FAIR Data Implementation in GGB Databases

	Based on your knowledge of and experience with FAIR data and your GGB database, please rate how much you agree with these statements:	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
C1	Standard nomenclatures are used consistently by the project(s) I work for, making it easier for researchers to find relevant data.	3 %	8	-	11	34	45
C2	Metadata is used consistently by the project(s) I work for.	3	3	5	8	47	34
C3	Common file formats are used consistently by the project(s) I work for to make it easier to share and integrate data from different sources.	3	3	-	3	50	42
C4	Tools and processes now used for contributing data to my project(s) are easy to use and facilitate accurate, efficient, thorough uploading of needed data and metadata according to FAIR principles.		5	8	29	42	16
C5	Tools and processes for finding and retrieving data from my project(s) are currently easy to use and facilitate accurate, efficient transfer of all needed data and metadata according to FAIR principles.		3	8	24	50	16
C6	The project(s) I work for provide good guidelines on what metadata to provide when preparing/submitting data.	3		5	24	38	30
C7	The project(s) I work for provide good guidelines on how to provide information about how the data have been generated and cleaned, how missing data has been handled, and other process transparency issues.	8	3	8	22	43	16
C8	The project(s) I work for provide useful resources for learning how to use them: tutorials, FAQs, how-to documents and videos, etc.	3	3	8	26	39	21
C9	The project(s) I work for highlight the importance of FAIR data principles for researchers and provides educational resources to help researchers understand and follow FAIR practices.	3	-	11	34	29	24
C10	The project(s) I work for employ cost-sharing efficiencies such as reusable open-source software.	-	-	3	11	34	53
C11	At this point in time, reliable interconnectedness and interoperability between related databases make it easy to combine and integrate data in new ways to address new questions.	5	8	21	32	21	13

Note. N = 37 or 38. Each row of numbers contains the proportion (percentage) of participants who gave each response to the question. Row percentages may not add to exactly 100% due to rounding. These questions were also asked in the 2022 baseline survey using identical or near-identical wording and in a different order; not all questions from 2022 were repeated in 2024.

2024 Priorities for Further Development of FAIR Data Practices in GGB Databases

Survey participants were asked to rate the importance of three potential priorities for improved data curation; their responses are summarized in Table 6. All three were rated as being "very important" or "highest priority" by more than half of respondents.

As in 2022, the highest ratings were given to the application of community standards by publishers and enforcement of data submission requirements," which was rated as "highest priority" by 54 percent of 2024 respondents and as "a very important priority" by another 43 percent of respondents.

In 2024, "Timely and up-to-date availability of curated data" was rated as being "very important" or "highest priority" by 84% of respondents. "Professional incentives from funders" was rated as being "very important" or "highest priority" by 73% of respondents. Corresponding figures from 2022 were similar (79% and 67%) indicating a sustained need for further development in these areas.

Additional comments on these priorities are listed verbatim in Table 7.

Table 6. 2024 Priorities for Further Development of FAIR Data Practices in GGB Databases

Please indicate what you believe to be the most pressing priorities for improving FAIR data compliance:		Not a Priority for Funding or Development	A Minor Priority	A Somewhat Important Priority	A Very Important Priority	Highest Priority for Funding or Development
D1	Timely and up-to-date availability of curated data	-	8 %	8	62	22
D2	Application of community standards by publishers and enforcement of data submission requirements	-	-	3	43	54
D3	Professional incentives from funders (e.g. enforcement of compliance with DMPs, credit for compliance with DMPs, inclusion of metrics for data sharing in evaluation)	-	8	19	27	46

Note. N = 37. Each row of numbers contains the proportion (percentage) of participants who gave each response to the question. Row percentages may not add to exactly 100% due to rounding.

Table 7. 2024 Comments on Priorities for Further Development of FAIR Data Practices in GGB Databases

Please add any other comments, observations or recommendations you'd like to share about the development of genomic, genetic, and breeding databases for shared research community use:
▪ *More sustained federal funding.* ▪ *Knowledge of different tools and software is crucial.* ▪ *Money or publications are the only incentives for researchers to take on the additional overhead of FAIR data.* ▪ *All of the above are very important, but until funders enforce data sharing requirements, the other two points are not as applicable. Having good curated data continues to be a highly under-appreciated, under-funded activity. Raising awareness of the importance of manually curated data to leverage funded research projects and not to prevent loss of important research results is still also a very, very important task that contributes to making data FAIR.*

Team Member Reports of AgBioData Impact on Participants and Their GGB Databases and Resources

In 2024 a new set of questions asked participants "In the past 2 years, how have you engaged with the AgBioData Consortium?" Respondents could choose more than one option. All responding team members reported participating in at least one AgBioData event or activity; 84% reported having responded to a previous AgBioData survey, 79% reported attending an AgBioData webinar, 71% reported attending an AgBioData workshop or presentation at a conference, 68% reported participating in an AgBioData Working Group, 63% reported attending an AgBioData community meeting (virtually or in person), and 58% reported having read one or more AgBioData papers.

Two new open-ended questions asked how participation in AgBioData impacted them personally or professionally and how this engagement could be made more valuable; see Tables 8 and 9 for their verbatim responses. In addition to potential useful individual comments in these tables, color coding has been applied to highlight several themes mentioned by multiple survey respondents:

- Collaboration and networking
- Improved knowledge and learning regarding standards, software, or other technical issues
- The need for more funding to support AgBioData participation.

As noted above in Table 4, 82% of participants "strongly" or "moderately" agreed that "AgBioData has helped me to become more familiar and experienced with FAIR data management principles and resources." Another 16% "slightly" agreed with this statement. One participant commented that *"I have learned a lot about FAIR and what it means and doesn't. I think most if not all scientists agree with FAIR principles but in fact need more funding and sustained database support to move forward. We need to make federally-funded databases a resource that is available, free to users and comprehensive."*

In 2024 an additional survey question asked respondents **"as a result of participation in AgBioData activities, has your database or related resource implemented (or made plans to implement) recommendations or tutorials or new technology for data sharing and management?" 10% of respondents answered "No" while 42% marked "Yes, we have implemented changes that are (at least partly) due to participation in AgBioData activities." Another 42% marked "Yes, we have plans to implement changes that are (at least partly) due to participation in AgBioData activities" and 16% marked "Not sure / N/A."**

Table 8. Comments on Personal and Professional Impact of AgBioData Participation

How has participation in AgBioData impacted you personally or professionally? For example, have you implemented any standards, tools or workflows in response to what you have learned? Have you initiated any new collaborations?
▪ *I have learned tremendously about ongoing databases use. I am retired from research so don't use databases directly anymore, but new collaborations have come from the work.*
▪ *It is nice to see there are other curators out there having the same issues as me. And the group has helped start the discussion about what to do and provided some solutions.*
▪ *Yes, through AgBioData network, I was able to secure travel support to attend BrAPI hackathon 2024 in Germany and was able to develop an R Shiny app for Plant Genome visualization called PlantGenomeViz already deployed to GitHub.*
▪ *I am considering implementing the nomenclature recommendations or recommending them to others. AgBioData also has me thinking more about DEI in databases now, and thinking about improved machine readability for our data (although we have not implemented any changes).*
▪ *I have learned from the nomenclature group about standards and proposed standards.*
▪ *I had been a very early participant, and my group is very active in AgBioData. The standards have been a huge benefit for us.*
▪ *I wish I spent more time working with AgBioData. Essentially, I use it as an information resource to know what all is going on. I'm not sure it has changed my behavior though I very much support its mission.*
▪ *Gained some new collaborations, learned of some new software.*
▪ *Yes, I have used participation in workshops to collaborate to get variation data into EVA.*
▪ *Networking.*
▪ *It has provided information on a number of data sources and opportunities they could provide in making the collection I curate more valuable to the research community.*
▪ *Some software we have developed has been adopted by other groups involved in AgBioData. Material presented at AgBioData webinars has impacted our project's use of some tools.*
▪ *I set up standards for cotton trait ontology.*
▪ *Community networking opportunities have made the biggest impact for me.*
▪ *Helped with ideas for presenting data, established collaborations.*
▪ *I have worked on G2P and Ontology working groups and currently initiated new working groups on sc RNAseq biocuration. We have collaboration with several labs to resolve the current challenges associated with gene expression metadata standards, tools and workflows.*
▪ *I have developed better capacity for writing research protocols and established collaborations with colleagues in my working groups.*
▪ *I have had the opportunity of scientific collaboration.*
▪ *Submitted 3rd party variant data to EVA to have RS numbers for a majority of variant sites in soybean.*
▪ *Participation in AgBioData was very good for me and for my profession, and I initiated new collaborations in new research projects.*
▪ *It enhanced my research knowledge, skills and experience.*
▪ *I have not worked much with AgBioData over the last 2 years but I did in the years before.*
▪ *I am more aware of the standards and how other databases have implemented them. Our databases do follow many aspects of the standards that are discussed in the AgBioData community.*
▪ *Networking.*
▪ *The booths at PAG and ASPB were most helpful in building community.*
▪ *Through AgBioData, I discovered a network of agricultural databases and found a wonderful medium to discuss and collaborate on issues such as data standards and application of FAIR principles.*
▪ *I have met new potential collaborators, I've learned new perspectives on data management, and in particular have learned how similar the challenges and solutions for data are between the plant and animal communities. We have not implemented new standards/tools but they have certainly informed our decision making process on standards and tools (i.e. in how to track metadata and what types of metadata to track). Yes, we initiated new collaborations.*

Table 9. Recommendations on Improving the Value of AgBioData Participation

What would make your engagement in AgBioData more valuable to yourself / your project?
▪ *More funding.*
▪ *All the work being done is great, but papers with the data not being available/deposited is still a major issue. Educating upcoming scientists and making recommendations to the community is important, but is not fixing the issue at hand.*
▪ *Currently working with the education working group has given the opportunity to interact with some already-established researchers in the field, and I have understood the need for Open Science and collaboration.*
▪ *More well-defined standards, format validation tools, and possibly some SOPs, best practices would be helpful.*
▪ *I've been pretty overextended. I don't think of anything that's likely to increase my engagement.*
▪ *Funds for time spent.*
▪ *More time or staffing to follow up on opportunities.*
▪ *More engagement with end users of database resources.*
▪ *I want to learn how people to standardize gene names.*
▪ *AgBioData consortium is a very good platform that facilitates community-based standards to maximize the accessibility and reuse of large scale data in agricultural research. It also provides networking opportunities to connect to 44 GGB databases and their members to resolve complex issues on metadata standards and FAIR data sharing.*
▪ *Enhanced project design capacity.*
▪ *AgBioData granted some exposure on scientific research publication.*
▪ *I would like to receive more newsletters about AgBioData research activities and trends in research about climate change.*
▪ *100 percent valuable.*
▪ *Win-win. AgroPortal (our ontology database) can support other databases.*
▪ *Broader global engagement.*
▪ *More frequent in-person meetings.*
▪ *It would be great to be able to have advisors from AgBioData to brainstorm on grants or to serve as advisors for grant proposals and to help provide letters of support.*