





AgBioData



Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

- Self-assessment of the long term financial stability of member databases
- Conduct a detailed analysis and modeling of sustainability solutions for representative AgBioData member databases.
- 3. Develop a roadmap for Genomic, Genetic and Breeding (GGB)
 Database sustainability to ensure data persistence and resource longevity (Years 2, 3, 4)





Self-assessment of the long term financial stability of member databases

Goals:

- Gather data through written surveys and interviews with staff at all member databases.
- Capture cost of operations, staff level, sources of funding, usage level, data types, species and strains, stakeholders served and anticipated future needs.
- Collect information on each GGB Database's view of its sustainability and approaches to improve that sustainability.
- Understand the current funding situation and the anticipated future needs.



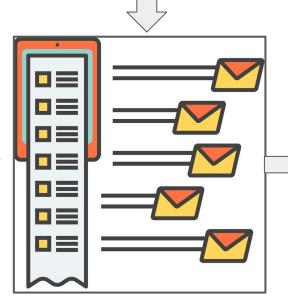
PI surveys



























AgBioData database Pls







Goal 1: Identify stakeholders



Stakeholder surveys

Identifying all stakeholder groups.

Recognizing the value that the database represents to stakeholders.

Understanding stakeholder attitudes toward sustainability models.



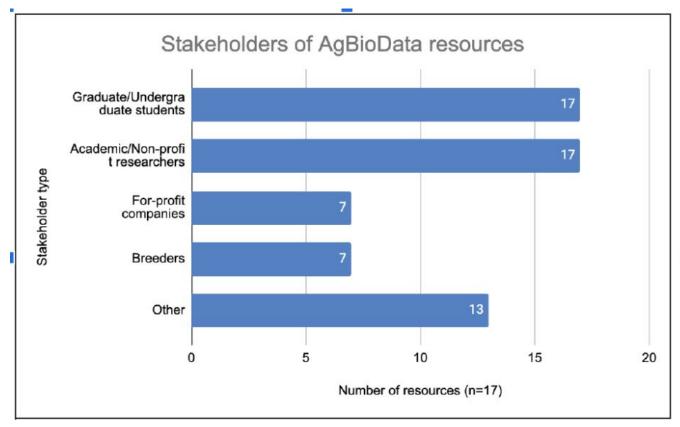


PI survey results





Stakeholders of AgBioData resources





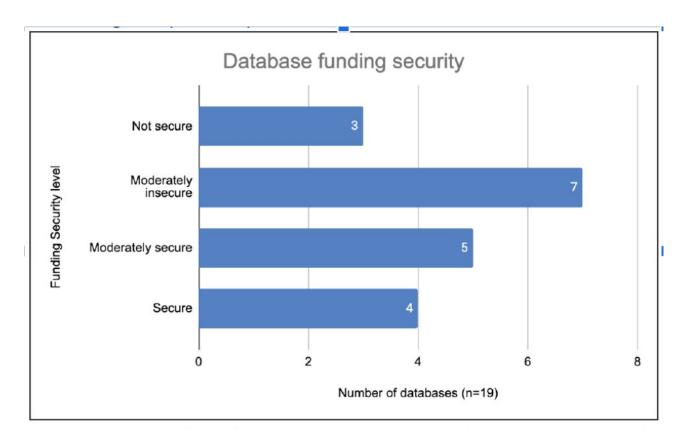




Goal 2: Urgency of action

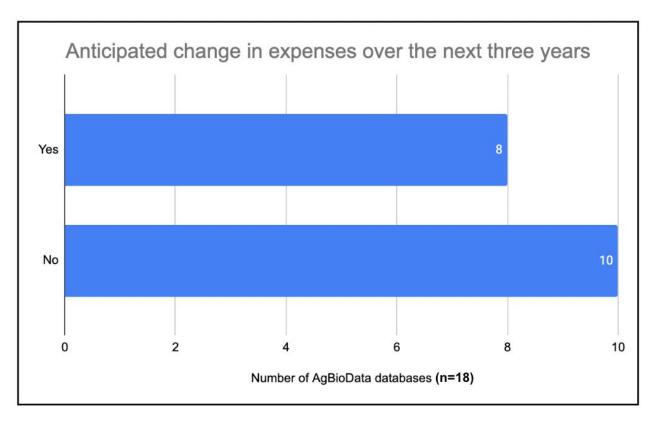


Current level of funding security



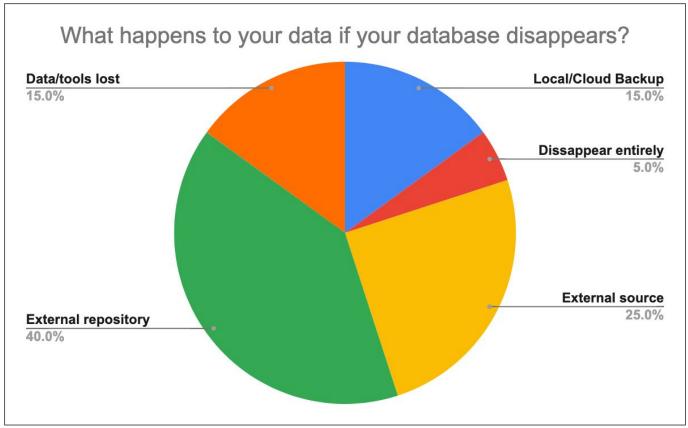


Anticipated change in expenses over next 3 years





Data persistence









Goal 3:
Database-specific facts/stats



What more do we need?

Database-specific facts/stats

Database size

- How much resources are needed to hold the data?
- How can the data be organized/distributed more efficiently?
- What opportunities are there to consolidate the data between databases?

Rate of data growth

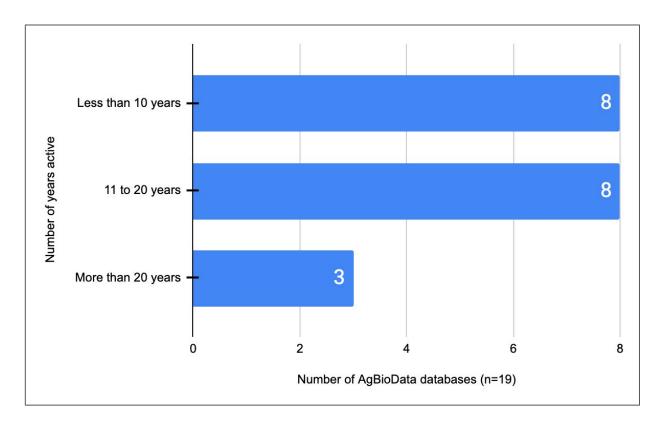
 How much resource allocation would be needed to sustain the current rate of data growth?

Data content

- What are the types of datasets within each database?
- What, if any, are opportunities for shared data content between databases?
- What are new datasets/types that would add value to the database?
- How much of your data is public?
- How much of it is curated in-house?
- How much of the data is imported from other resources?

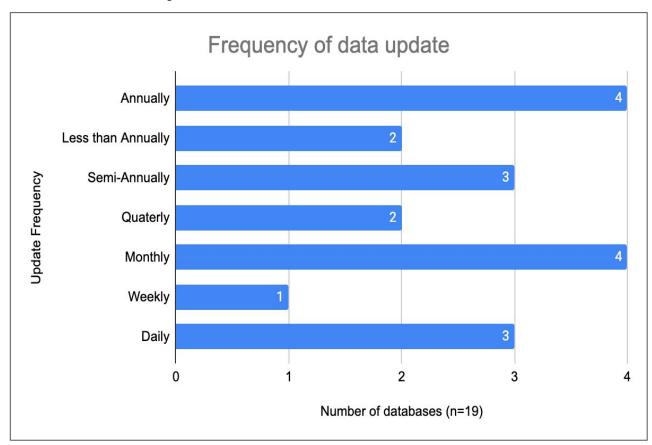


Number of years active





Frequency of data updates





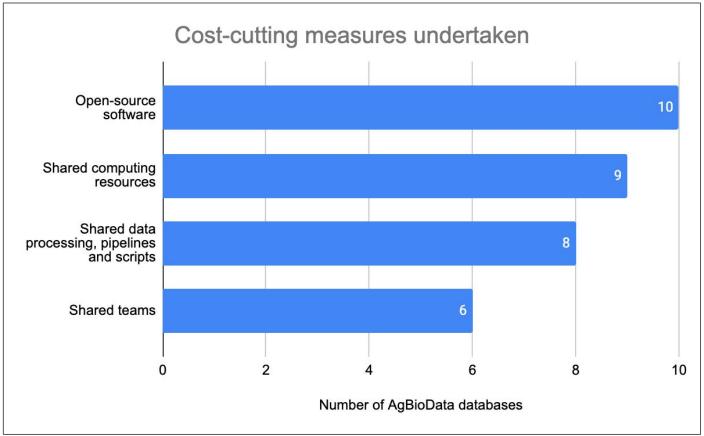




Goal 4: Cost-sharing strategies

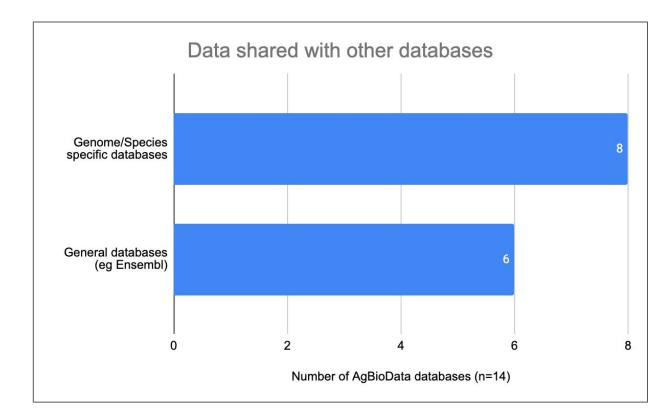


Steps to reduce expenses





Data shared with other databases





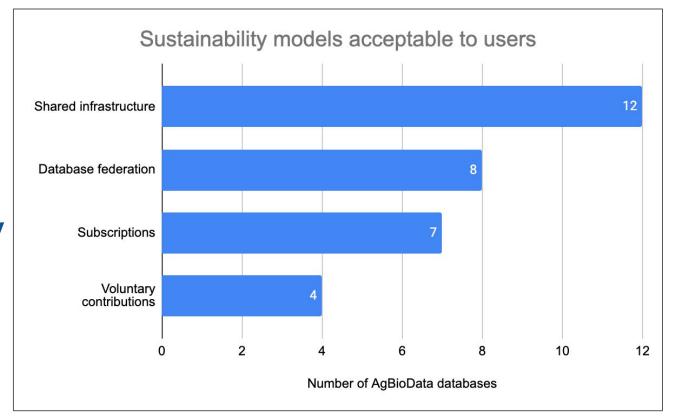




Goal 5: User experience

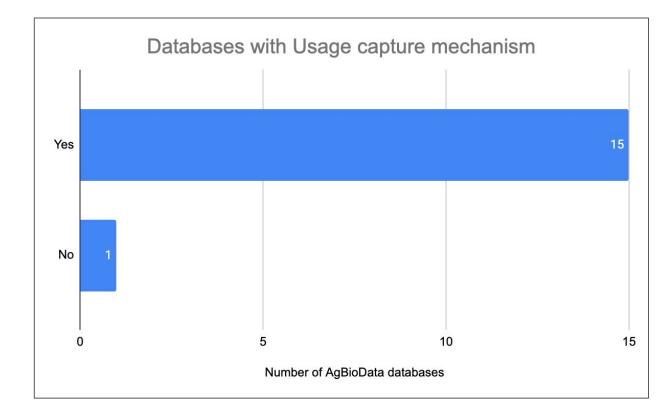


Acceptable sustainability models



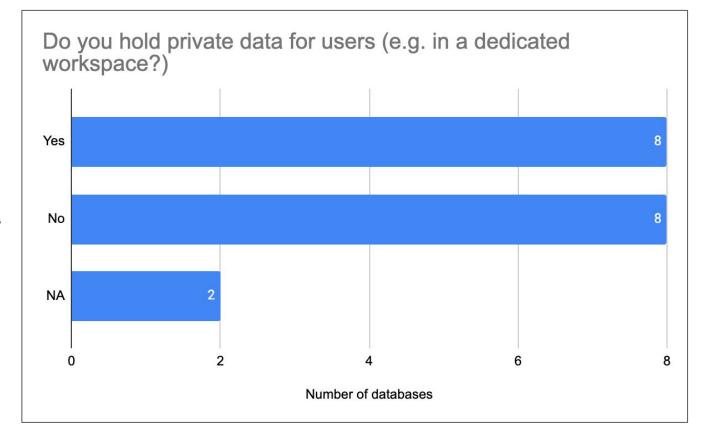


Usage capture mechanism





Private data workspace







What does the survey tell us?

- Never too early to plan for a sustainable future. 50:50 split in level of funding security. So it's never too early to look into sustainable options.
- For a majority of databases the curated part of their data will no longer be available if the database disappeared completely. So sustainability solutions need to be considered with curated datasets in mind. Curated datasets are a great add in value to your database. So the more curated datasets you can add to your databases the better.





Recommendations

- Mechanism to capture usage statistics: If your database does not have a usage capture mechanism, implement one.
- Have an authentication system. This allows users to store private data, workflows, and analysis.
 Shared authentication system
- Database participation is key.
 - We need databases to be more responsive to our surveys.
 - The more data we have the more we can develop a solution that works for all AgBloData databases as a group. Model depends on data.
 - If you prefer one on one conversations like an interview style, we will work with you on that.







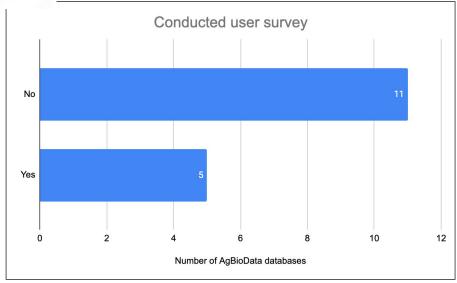


User surveys

- Users: academic and commercial
- Do you have any user surveys completed in the last 2 years?
 - Important to understand if the user mindset has changed.

Use Cases

- How do stakeholders utilize the data/tools within the database?
- Where are opportunities for shared use cases?
- Surveys of other stakeholders including:
 - Employees
 - Crop commissions
 - Funding agencies
- Where are untapped opportunities for the data/tools within the databases?









1. Funding models:

- a. Voluntary membership
- b. Data deposit fees
- c. Subscriptions
- d. Freemium models
- e. Donations



Voluntary Membership





2. Cost sharing:

- a. Shared curation or software teams
- b. Community Curation
- Shared infrastructure







Community curation





Goals of the Sustainability Working Group

1. New funding sources.

- a. Assess the viability of various sustainability models
- b. Recommend sustainability models for AgBioData databases, either as a consortium or for individual databases.

2. Cost-saving measures.

- a. Determine effective strategies for cost sharing including sharing data storage, FTEs, Tools etc.
- b. Evaluate the idea of a centralized approach to resource management benefitting all member databases.

3. Policy or procedure changes related to grant funding.

- a. Increase the ability of databases to obtain a share of research funding that supports data generation.
- b. Examine the role that a central coordinating entity could play in assisting with sustainability.

Acknowledgements



RCN NSF Award 2126334

AgBioData Steering committee

Annarita Marrano Leonore Reiser Monica Poelchau Meg Staton Jaqueline Campbell Peter Harrison Sook Jung Sunita Kumari John McNamara Sushma Naithani

Director.

Phoenix Bioinformatics Family



Tech Team





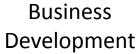
Sales





Finance







Science Team













Swapnik





Contact information and PAG30 schedule

YouTube : @phoenixbioinformatics8986

Facebook : phoenixbioinformatics

Twitter : @phoenix_bio

⊠ Email : shabari<u>@phxbio.org</u>

Join us for our workshop on sustainability:

Data Resource Sustainability and Funding Monday, Jan 16 4:00 PM @ Palm 7







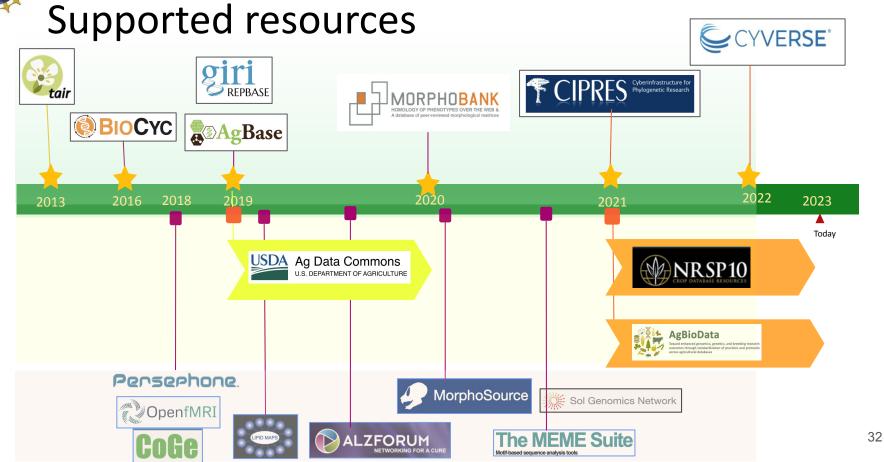




NSE

Partners, Sustainability consulting and





Principal Investigator survey goals



Identify stakeholders

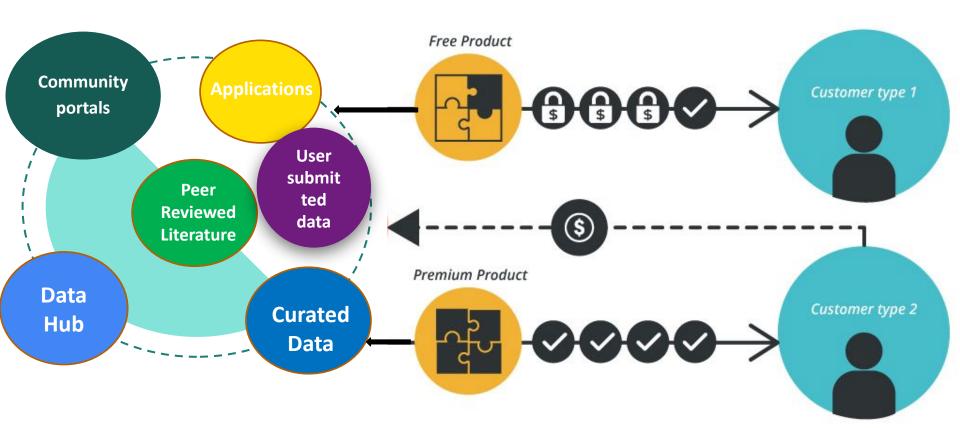




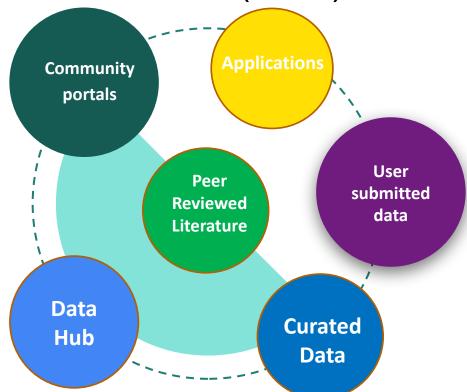




Freemium

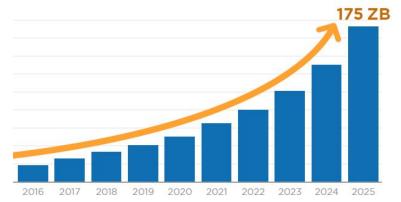


Digital Data Resource (DDR)





Biological data resource sustainability



Volume of data from scientific research is rapidly increasing

Data Volume + Increased requirements for access= Increased resources for data storage and analysis



CATALOG Assign metadata schema.

specialized and disciplinary taxonomies, digital object identifier (DOI), and universal numerical fingerprint (UNF)

MANAGE Administrative online research data archives

Project data from

experiments, surveys,

researchers.

and scientists





SYNTHESIZE RESEARCH

Verification, insight, discovery, visualization, harvesting, and linked data

Digital data resources (DDRs allow for collaborative discovery

Challenges to Digital Data Resource Sustainability



FUNDING

- DDRs are growing rapidly while funding flatlines.
- Grants focus on funding new research at the expense of saving prior data.
- Heavy reliance on a single source of funding, which is not sustainable.

Data Management

- Lack of data standards, Increasing computing costs, Rapidly increasing volume of data
- Making current data available.





DDR community/ user culture

- Cultural resistance to data sharing within their user community.
- Misconceptions that computing is free.
- Data sharing and management is typically not valued for career advancement.

Unpredictable future

- Understanding and predicting future sustainability challenges is daunting
- Future data sharing policies and funding need buy in from all stakeholders.



Working groups











Public Genetic Resources

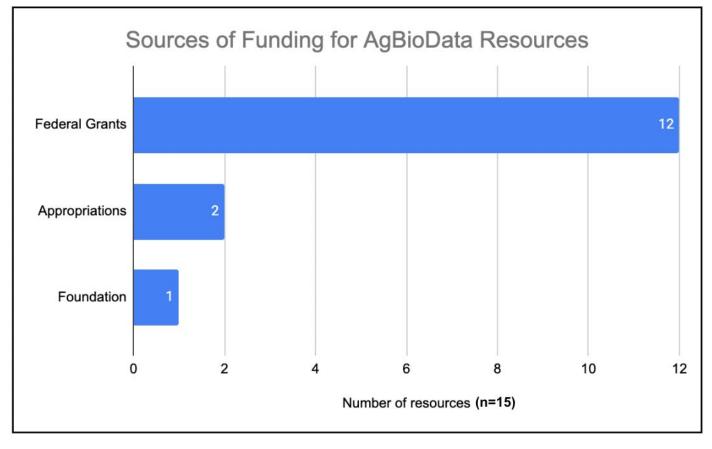


Data Federation Training



Urgency of action

Sources of funding for AgBioData databases





Data archived elsewhere

