Sustainability RoadMap for AgBioData Databases

The AgBioData Consortium Workshop
Jan 10, 2025

Goals of the NSF RCN

1. Self-assessment of the long term financial stability of member databases
2. 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)

Aim 1. 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.

Stakeholder surveys

Identifying all stakeholder groups.

Recognizing the value that the database represents to stakeholders.

Understanding stakeholder attitudes toward sustainability models.

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 from PI surveys

- **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 AgBIoData 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.

Charge to Sustainability Working Group

RCN Aim 4.2: Conduct a detailed analysis and modeling of sustainability solutions for representative AgBioData member databases.

This includes:

- Exploration of New Funding Avenues
- Investigation of Cost Reduction Techniques
- Assessment of Policy or Procedural Revisions Pertaining to Grant Funding

Exploration of New Funding Avenues

- Voluntary membership
- Subscription models
- Implementing fees for data deposit
- Corporate backing
- Philanthropic contributions
- 'Freemium' options
- Crowdfunding

Investigation of Cost Reduction Techniques

- Collaboration in Curation and Software Development
- Curation by the Community or Data Providers
- Utilizing Shared Infrastructures

Assessment of Policy or Procedural Revisions Pertaining to Grant Funding

- Mechanisms to Capture User Interest
- Document the Importance of the Database to Its Stakeholders
- Mechanism to Capture Citations
- Tracking the Global Core Biodata Resources

RCN Aim 4.3: Develop a roadmap for GGB Database sustainability to ensure data persistence and resource longevity: Key Findings

- Importance of Stable Long-Term Funding
- Community Value of AgBioData Resources
- Diversity of Funding Sources
- Enhancing Computational Resources for High-Throughput Omics and Genome Analysis
- Strengthening Community Engagement
- Identifying Commercialization Opportunities
- Investing in AI and Machine Learning

Recommendations

- Highlighting Resource Impact through Secondary Citations
- Addressing Staffing Challenges through Subcontracting
- Integrating AgBioData Resources into Large-Scale Project Grants
- Communicate the economic impact of the database

Acknowledgement

Thank you to the AgBioData Sustainability Working Group members for their valuable input and guidance.

This work is made possible by the support of the National Science Foundation through Award Abstract # 2126334.