AgBioData

NSF RCN: Reimagining a sustainable data network to accelerate agricultural research and discovery

https://www.agbiodata.org

Today's Agenda

▶ **Introduction to AgBioData** - Lisa Harper
▶ **The AgBioData RCN grant** - Eva Huala
▶ **Concept of the Working groups** - Monica Poelchau
  ▶ Genome Nomenclature - Lisa Harper
  ▶ Metadata Standards - Ethy Cannon
  ▶ Data federation - Monica Poelchau
  ▶ Emerging data types - Sook Jung
  ▶ Recruitment - Jacqueline Campbell
  ▶ Implementation - Jacqueline Campbell
  ▶ Education - Meg Staton
  ▶ Sustainability- Eva Huala
▶ **Opportunity to Join!** - Leonore Reiser
▶ **First Virtual Meeting - Mid- February 2022**
▶ **Conclusion**

Today's Agenda

Please put your questions in the chat box!

What is AgBioData?

Jan 2015, Scientists from about 15 Agriculturally related Genetic, Genomic and Breeding Databases got together at PAG to discuss how to work together to solve common problems in a consistent and systematic way.
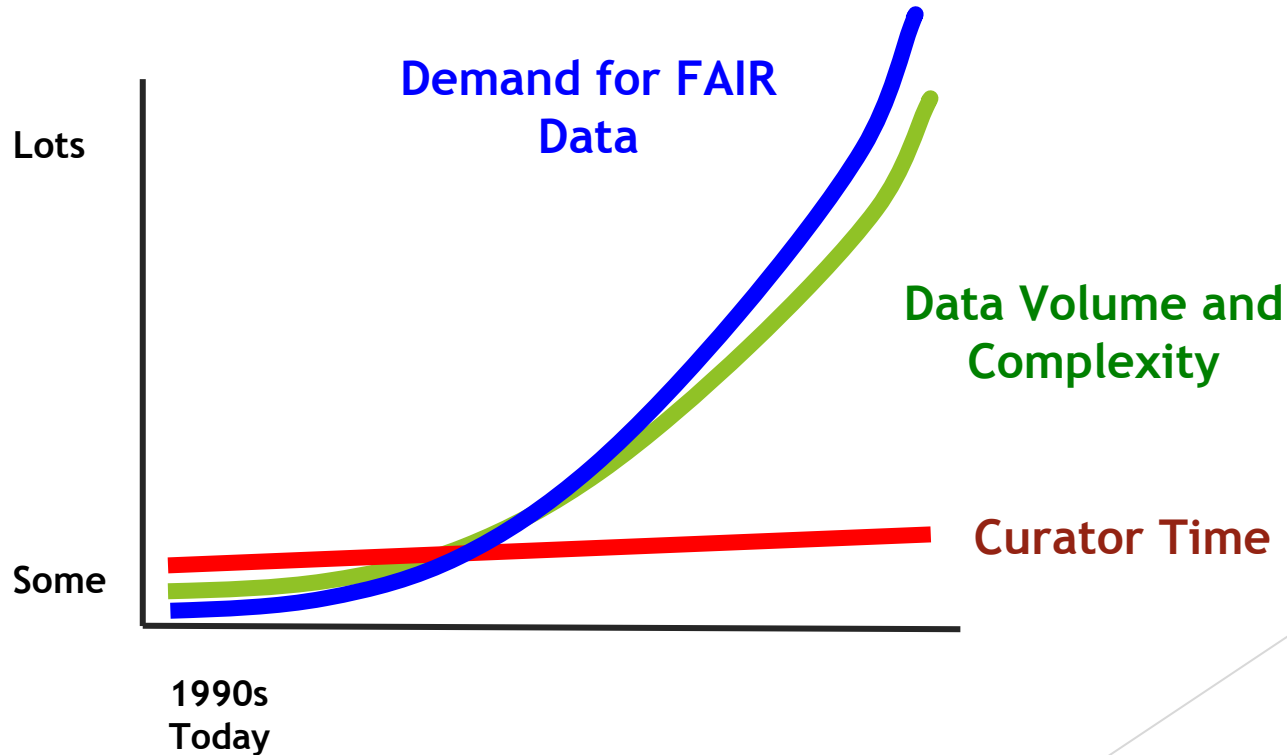
What is AgBioData?

▶ Freely available DATA is critical for scientific discovery and reproducibility.

▶ **Since the 1980s, Databases** store and allow access to data, and establish standards for data storage, descriptions, access and sharing.

▶ Data Curation has become increasingly important as data volume and complexity increases.

What is AgBioData?

Genetic, Genomic and Breeding Data

Demand for FAIR Data

Data Volume and Complexity

Curator Time

Lots

Some

1990s
Today

What is AgBioData?

If we could work together to solve common issues, we could leverage valuable curator effort, and create a more uniform user experience

▶ Established initial Working Groups

▶ Established monthly seminars, which continue to this day

▶ In person meeting, Salt Lake City 2017

▶ Top 7 priority issues established, along with new working groups

▶ Published and extensive white paper in 2018 (https://pubmed.ncbi.nlm.nih.gov/30239679/)

What is AgBioData?

>150 Scientists from >30 Databases and Resources

2021 Steering Committee

Lisa Harper* MaizeGDB

Monica Poelchau* NAL i5K

Jacqueline Campbell* SoyBase

Ethy Cannon* MaizeGDB

Sook Jung* GDR

Eva Huala Phoenix BioInformatics

Meg Staton Hardwood Genomics

Leonore Reiser TAIR

Laurel Cooper Planteome, GrainGenes

Marcela K. Tello-Ruiz Gramene

Darwin Campbell DILL-PICL Lab Iowa State Univ.

Really Important Super Members rotated off: Ramona Walls (CyVerse), Dorrie Main (GDR, CottenGEN, etc)

*   2016 Founding Members

What is AgBioData?

▶ **Join our mailing list**

▶ **Join a Working Group**

▶ **Participate in a webinar**

▶ **Connect via social media:**

@AgBioData

Slack
https://www.agbiodata.org/join-slack

The AgBioData RCN

What's an RCN?

Research **C**oordination **N**etwork -
Funded by National Science Foundation

… supporting groups of investigators to **communicate and coordinate** their research, training and educational activities across disciplinary, organizational, geographic and international boundaries.

Overarching RCN Goal:

Make agricultural genetics, genomics and breeding (GGB) data easier to find and use, and available for the long term

When: Sept 1 2021 - Aug 31 2024 (3 years)

Funding: includes a full time Program Coordinator,
in-person meeting in year 2, sustainability planning

AgBioData RCN Aims

1. Develop **recommendations, standards and implementation plans** for FAIR data

Findable
Accessible
Interoperable
Reusable

BEST PRACTICE

2. Expand the network

Funders

Databases

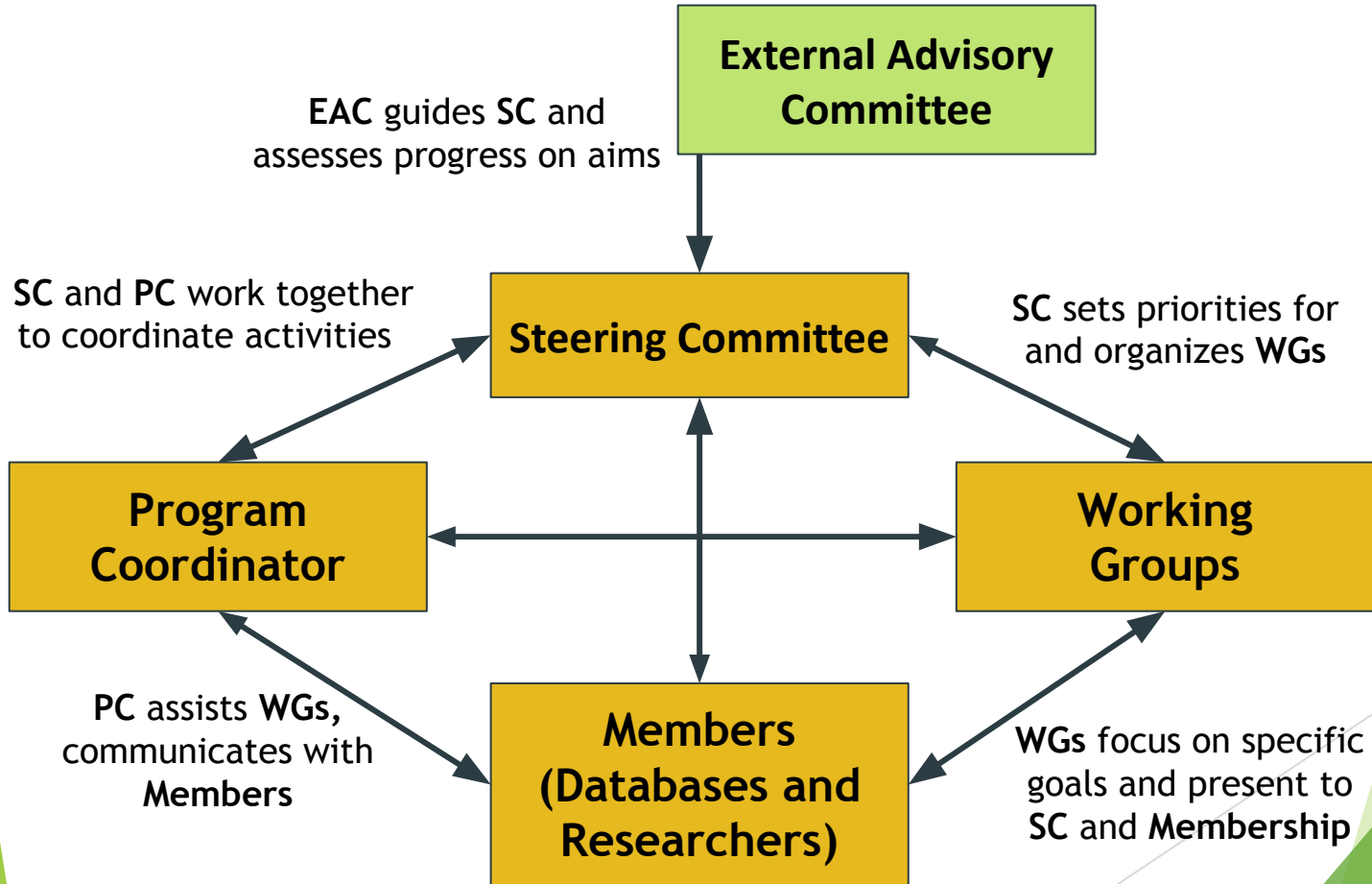Researchers

Publishers

3. Develop **educational materials** to train researchers about **FAIR** data standards

4. Develop a **roadmap for a sustainable GGB Database Ecosystem**

AgBioData Organization

External Advisory Committee

EAC guides **SC** and assesses progress on aims

Steering Committee

SC and **PC** work together to coordinate activities

SC sets priorities for and organizes **WGs**

Program Coordinator

Working Groups

PC assists **WGs**, communicates with **Members**

Members (Databases and Researchers)

WGs focus on specific goals and present to **SC** and **Membership**

AgBioData RCN workshops

- ▶ Three workshops planned during this RCN - one per year

- ▶ Purpose of the workshops - to come together to report on working group progress, and share and develop ideas

- ▶ Expected in February 2022, 2023, 2024

- ▶ 2-3 days each

- ▶ The 2023 workshop will be in-person

- ▶ Stay tuned for details!

AgBioData Working Groups

▶ A Working Group (WG) is a group of AgBioData RCN participants working regularly together to define and solve a problem
  ▶ 'Issue' Working groups - developing recommendations, standards and implementation plans for FAIR data (Aim 1)
  ▶ Focused working groups from Aims 2, 3, and 4 - recruitment, implementation, education, and sustainability
▶ AgBioData has already successfully implemented the WG model
  ▶ Example: The GFF3 working group

▶ The WGs are *central* to the success of this RCN - most of the work of this RCN will be handled in WGs.

Working group organization

► 1 volunteer chair and co-chair
  ► Can be a Steering Committee or general network member

► At least 6 WG members

► Includes established professionals and students or postdocs

► Recruitment working group will help with member diversity
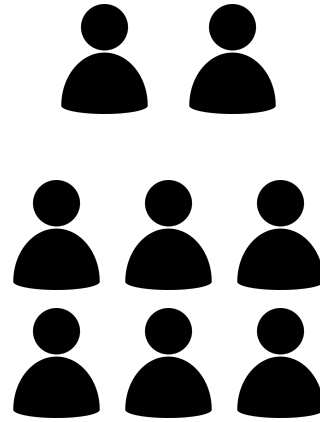
Working group time commitments

► Workshop attendance. Attend most or all of that year's workshop

► Regular online meetings. Meet at least once per month.

► WG membership duration
  ► Ideally one year
  ► 6 month term limits are feasible, in particular for students or post-docs who might be moving on to other positions

Working group activities

► WG chair, co-chair: report monthly progress to the Program coordinator; organize WG meetings; ensure that WG deliverables are met

► Work commitments. Contribute substantially to the WG with verbal and/or written contributions

► Deliverables. There are specific deliverables for each working group; however, in general:
  ► Establish a set of objectives to be accomplished within 6-12 months
  ► Develop an implementation plan to accomplish those objectives.
  ► Produce a final written report on the WG's findings

Timeline

https://time.graphics/line/540454
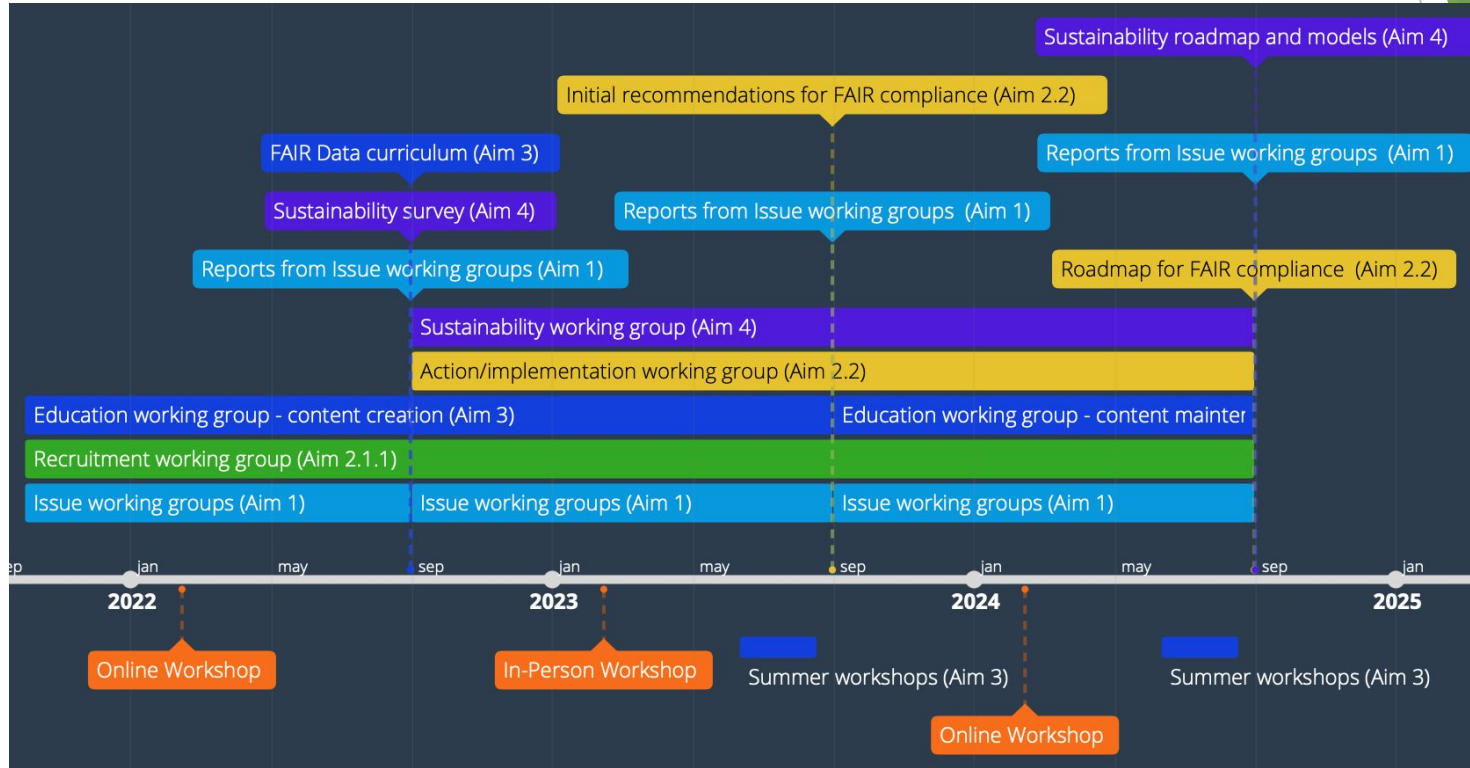
Working group assembly

▶ Poll at the end of this webinar will assess your interests

▶ Survey one week after this webinar will gather additional WG suggestions from you

▶ A second survey in mid-September will be sent out to gather participants, chairs and co-chairs for each WG

▶ Recruitment working group will ensure that students and postdocs are included in each working group

▶ Working groups should begin in October

Benefits of Working Group membership

► Co-authorship on working group reports

► Acknowledgement on AgBioData website as a working group member

► Network with colleagues and develop new collaborations - potential for new grants

► Students and post-docs are especially welcome - this is a great opportunity to expand your professional network

► Contribute towards better data management for your research community!

AgBioData Working Groups

▶ Genome Nomenclature - Lisa Harper

▶ Metadata Standards - Ethy Cannon

▶ Data Federation - Monica Poelchau

▶ Emerging data types - Sook Jung

▶ Recruitment - Jacqueline Campbell

▶ Implementation - Jacqueline Campbell

▶ Education - Meg Staton

▶ Sustainability- Eva Huala

Genome Nomenclature Working Group

▶ Reference quality genome assemblies of multiple cultivars of individual species are becoming available.

▶ Each assembly can have multiple versions and multiple annotation sets.

▶ Individual gene models need to be tied to annotation sets and specific assemblies.

This is an emerging issue. Now is the time to establish a method to insure globally unique, persistent identifiers for each assemblies and annotation set.

https://youtu.be/kNW6YReFP28- Genome Nomenclature
https://www.youtube.com/watch?v=m-Q0wn_Vgtg- Dealing with 50 genomes

Metadata Working Group

▶ Long-standing WG that has explored existing metadata standards for a variety of data types.

▶ Mission is to improve use of standardized metadata, which is critical especially for finding and relevant datasets and sharing and reuse of data.

▶ Proposed activities:

    ▶ Document recommended metadata standards for member databases.

    ▶ Develop a metadata assessment tool.

    ▶ Work with member databases to document metadata practices.

Data Federation Working Group

► Data federation: the ability to discover and aggregate data from various sources

► Data federation is a top priority to mitigate data discovery and exchange problems (Harper *et al*. 2018)

► Previously identified problems:
  ► AgBioData data is distributed across many databases;
  ► Data federation across GGB Databases is not standardized.

► Proposed Year 1 focus for AgBioData RCN:
  ► Identify challenges to data federation due to diversity of storage systems and data types.
 ► Define a federated data exchange model for one (+) data type across multiple AgBioData databases
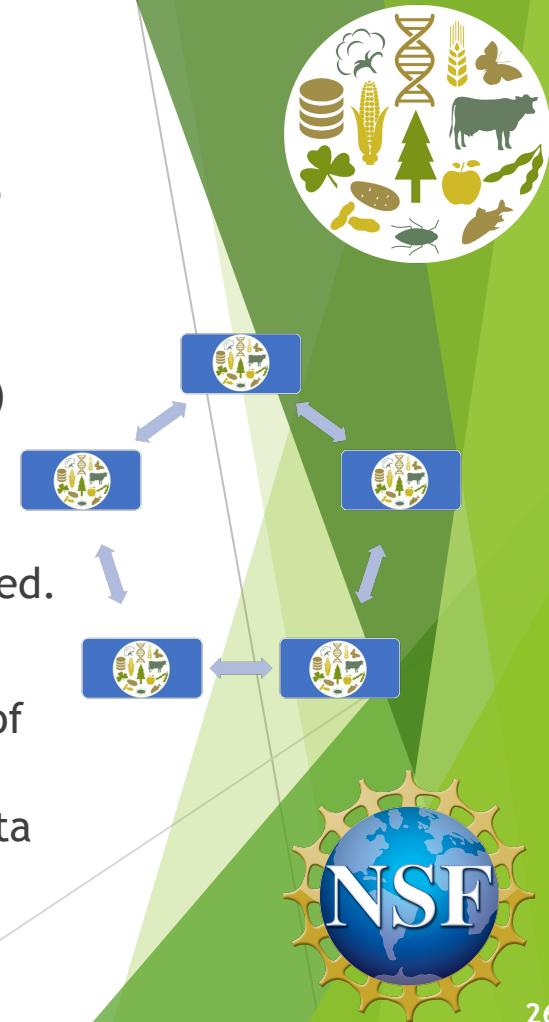
Emerging data types Working Group

► Novel data types often results from novel scientific methods

► The goal of this working group is
  ► To select the most relevant new data types for AgBioData members
  ► Identify the specific challenges to make this data type FAIR
  ► Pass results onto other working groups (eg. metadata, data federation) in following years to develop best practices
  ► We could form a new issue WG after 6 month

► Example new data types
  ► High-throughput phenotyping data
  ► Metabolomic data
  ► Image data
  ► Microbiome data

Recruitment Working Group

Funders · Databases · Researchers · Publishers

▶ **Aim:** Expand the network to include key stakeholders

▶ **Goal:** Broadening participation in the AgBioData network

▶ Focus efforts to recruit & engage participants
  ▶ Underrepresented groups (NIFA Tribal Colleges)
  ▶ Junior researchers (graduate students/postdocs)

▶ Identify relevant groups/organizations, establishing contacts, & gathering/reporting on metrics

▶ WG will meet virtually & at the beginning of each year
  ▶ To recruit data generating scientists &/or stakeholders for each working group (Program Coordinator will play a role in this work)
  ▶ End of year 2, recruit people from agricultural scholarly publishers, major federal & private funding agencies to join the **Implementation Working Group**
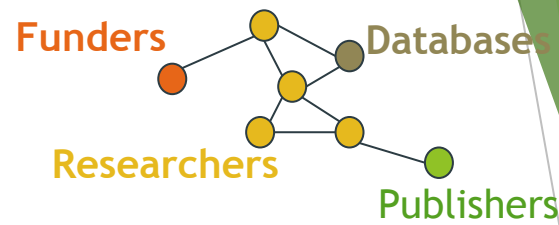
Implementation Working Group

▶ Starts during Q1 of Year 2, with AgBioData members, stakeholders, data generators, agricultural scholarly publishers, major federal & private funding agencies

▶ Working group will implement community-wide Aim 1 deliverables and develop effective mechanisms

  ▶ Develop effective mechanisms and strategies to implement the FAIR data management

  ▶ Publish an initial set of recommendations/roadmap for funders and journals to support and encourage FAIR data compliance.

Education Working Group

- ▶ FAIR Data Management Guide
  - ▶ Researcher-focused
  - ▶ What GGB databases are out there? Which should I use?
  - ▶ Where and how should I submit my data of type X?
  - ▶ Checklist of FAIR data procedures
  - ▶ Start general, then add specifics by data type as time allows

- ▶ FAIR data curriculum
  - ▶ Foundational knowledge of FAIR and GGBs
  - ▶ Graduate student courses
  - ▶ Workshops at Conferences
  - ▶ Online, at your own pace

- ▶ FAIR data summer workshop
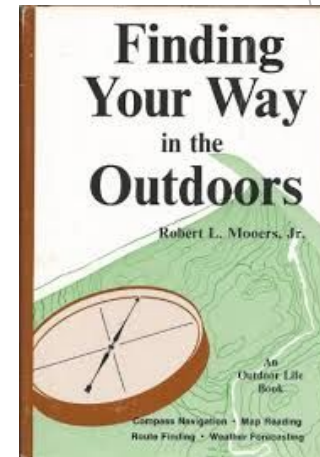
Sustainability Working Group

- Assess financial sustainability of AgBioData member databases
  - Survey 34+ member databases on financial situation
  - Make results available on AgBioData website

- Explore sustainability options for a few member dbs to serve as examples

- Develop Sustainability Roadmap for the consortium and its member databases
  - Individual database support models
  - Reduced costs through shared resources
  - Changes to funder policies to improve db sustainability

POLL: What working group(s) are you interested in...?

► Nomenclature -

► Metadata Standards -

► Data federation -

► Emerging data types -

► Recruiting -

► Implementation -

► Education -

► Sustainability -

Leonore Reiser Ph.D.
TAIR

What other working groups do we need?

Add your suggestions to the chat

Save the date for our first working meeting!

AgBioData Virtual Workshop

Mid-February (around Feb 14-28, 2022)

Love your (AgBio) Data Week

Next steps

► Complete the follow up surveys. Your input is ESSENTIAL, and we appreciate your taking additional time for this.

► Join your working group (starting October 2021)

► Attend the virtual workshop in Feb 2022

► Stay connected
  ► Join AgBioData/subscribe to email updates
      https://www.agbiodata.org/user/register

► Follow us on social media

  📧 @AgBioData

  Join us on Slack
  https://www.agbiodata.org/join-slack

Questions?