

# FAIR Scientific Literature Working Group

AgBioData Community Workshop  
5/1/2023



# FAIR Scientific Literature WG 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.
- [Second year] publish recommendations and a roadmap for authors and publishers to increase the FAIRness of published research.



# Members



Ruth Isaacson  
GSA



Leyla Cabugos  
Librarian, Cal Poly



Jenna Daenzer  
GSA



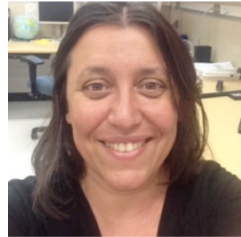
Leonore Reiser  
TAIR curator



Sook Jung  
Asst Research  
Professor. GDR



David Molik  
Computational  
Biologist,  
USDA ARS



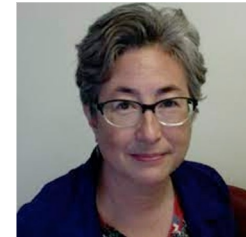
Daniela Raciti  
Exec Editor,  
microPublication &  
Wormbase Curator



Jacqueline Campbell  
Geneticist, USDA ARS  
SoyBase curator



Adam Wright  
Software Engineer  
Wormbase.Reactome



Karen Yook  
Exec Editor,  
microPublication &  
Wormbase Curator



Daniel Morris , Professor (advisory)



# Genetic, Genomic & Breeding (GGB)

**COMMUNITY GGB  
CURATOR**

**(Make Data Findable, Accessible,  
Interoperable, Reusable)**

# Current publication-curation challenges

**AUTHOR**  
Data Maker

**Databases**



# Current publication-curation challenges

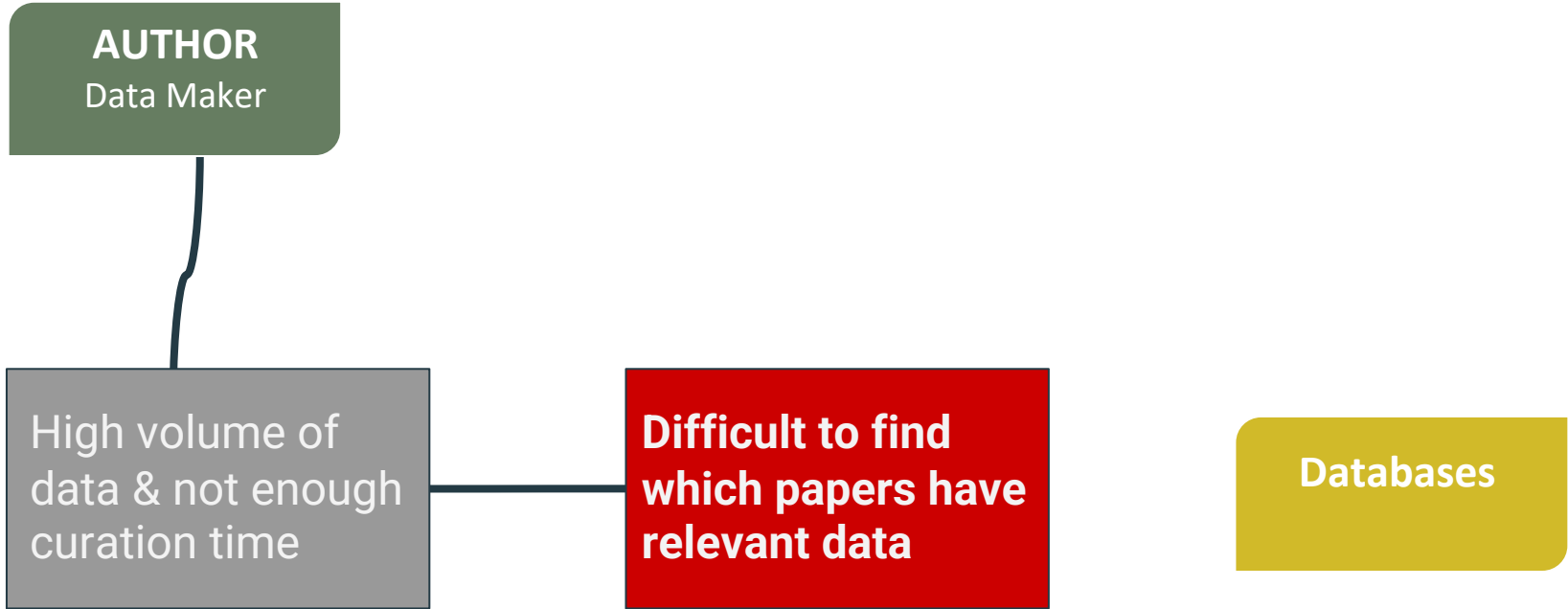
**AUTHOR**  
Data Maker

**High volume of  
data & not enough  
curation time**

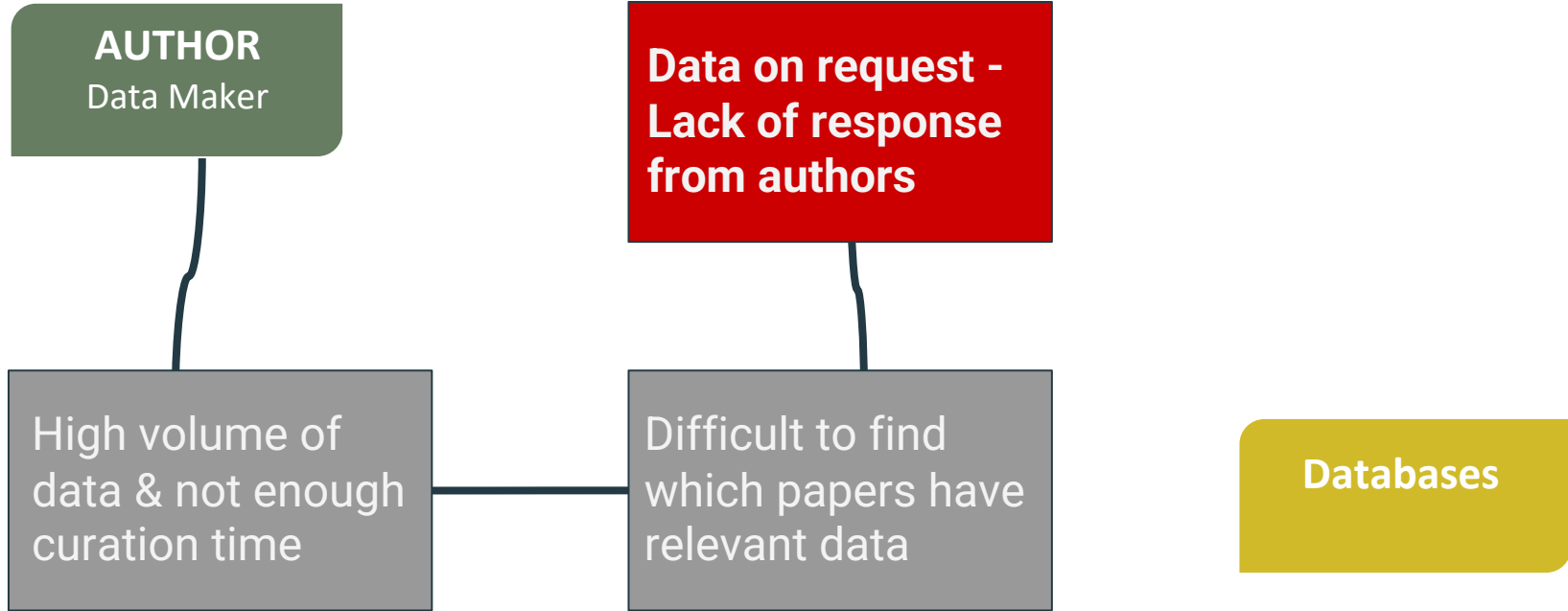
**Databases**



# Current publication-curation challenges

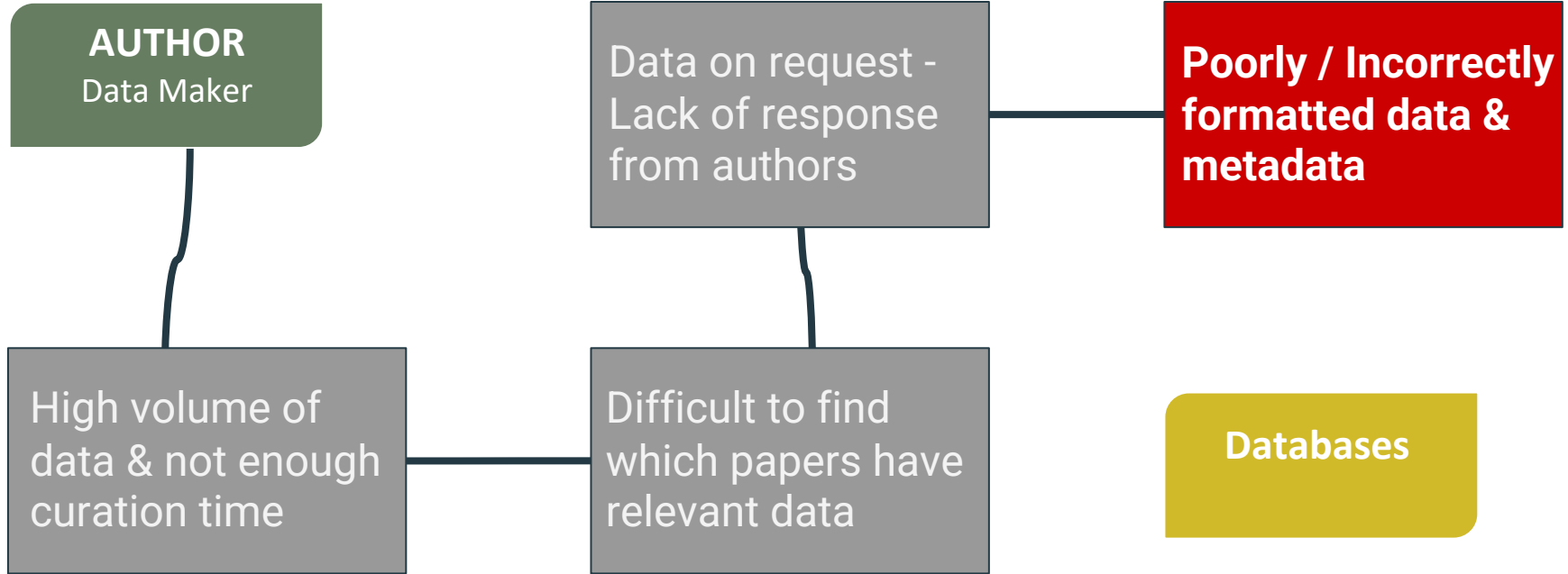


# Current publication-curation challenges

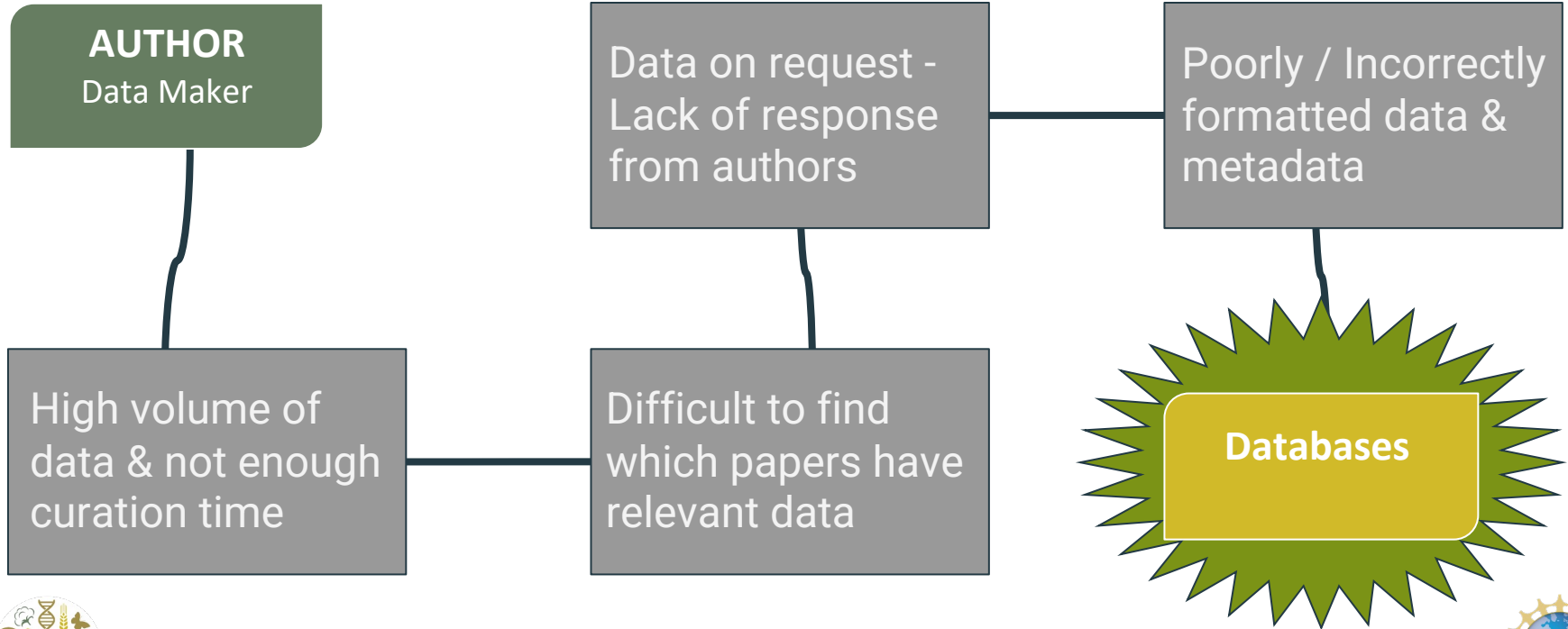




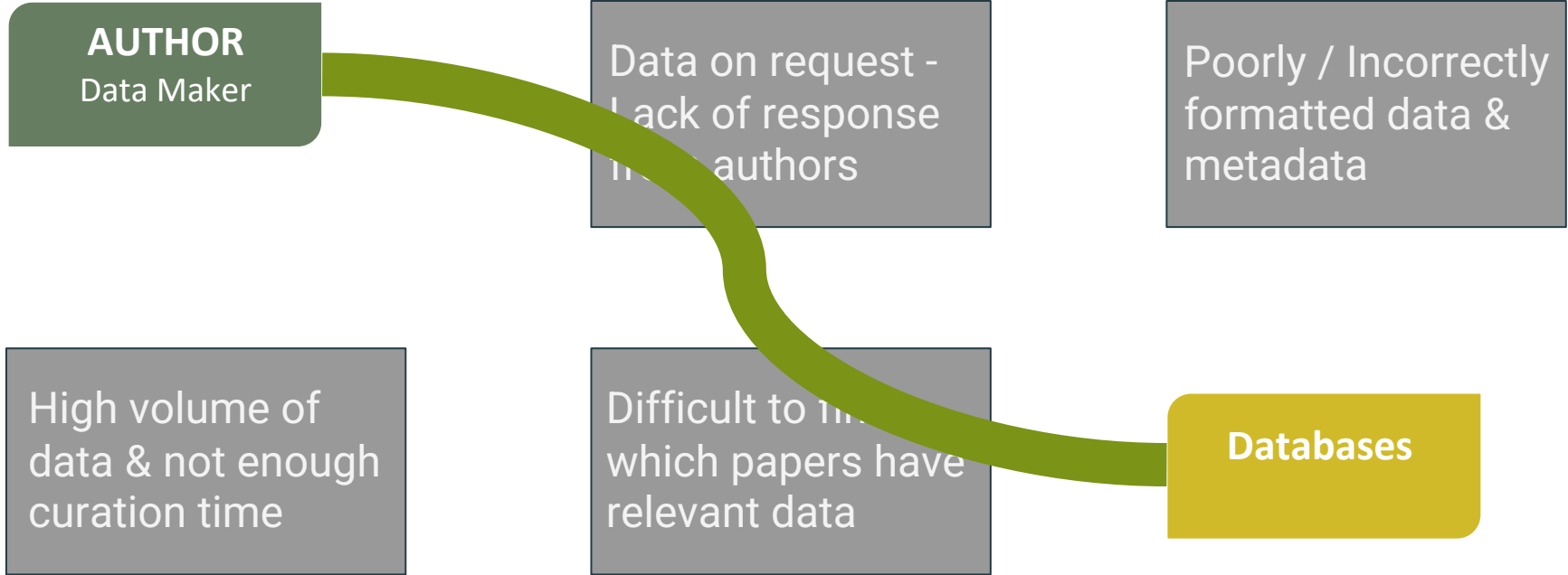
# Current publication-curation challenges



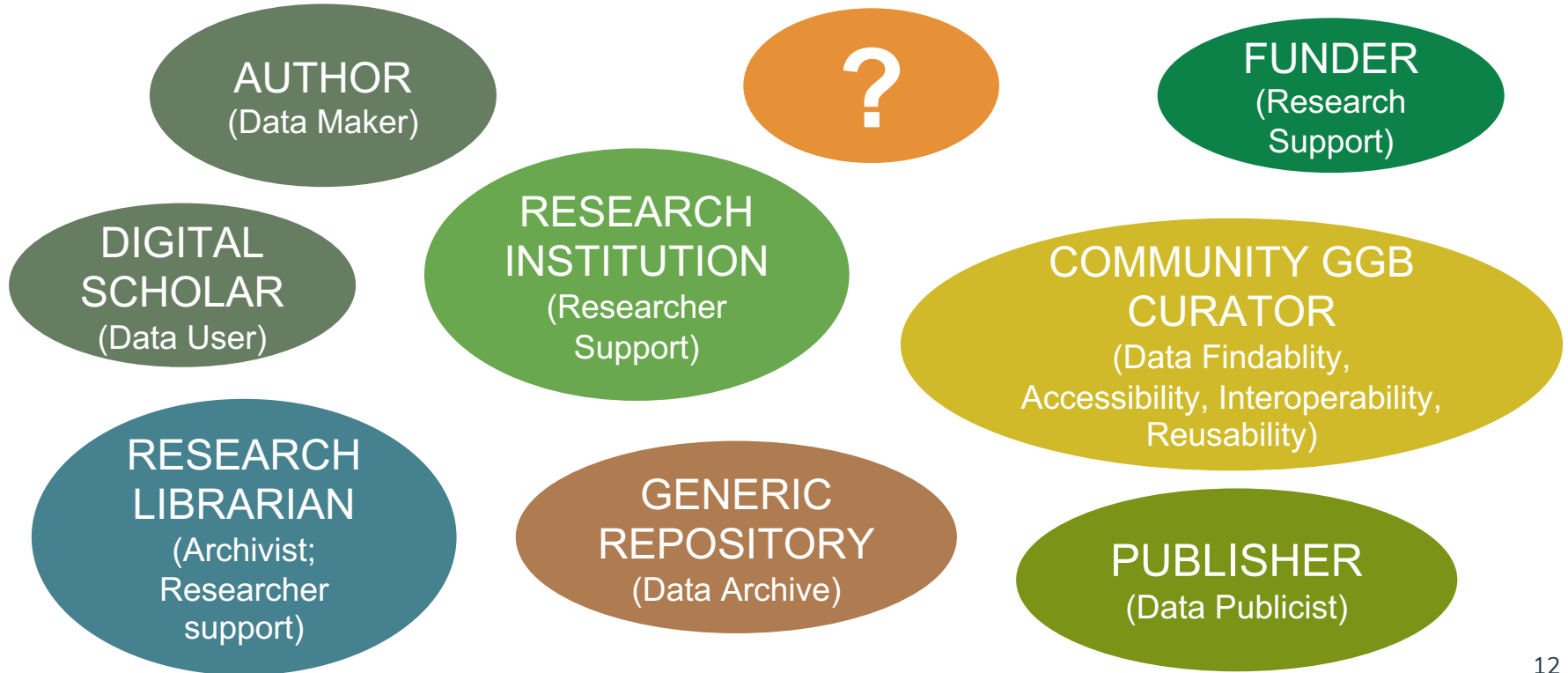
# Current publication-curation workflow



# Aim for a BETTER workflow



# Stakeholders involved w/ data management



# Focus on Four Stakeholders

**AUTHOR**  
(Data Maker)

**FUNDER**  
(Research  
Support)

**RESEARCH  
LIBRARIAN**  
(Archivist;  
Researcher  
support)

**PUBLISHER**  
(Data Publicist)

# Four Stakeholder challenges & barriers

## Researcher

- **Not knowing where data should go**
- Time consuming submission process
- Not knowing how to format data / metadata

## Publisher

- **Not knowing where data should go**
- Lack of easy verification of data availability
- Authors often do not want data available before publication

## Librarian

- **Not knowing where data should go**
- Lack of familiarity with data
- Lack of engagement with researcher

## Funder

- **Not knowing where data should go**
- Different programs have different repositories
- Proposal reviews don't know how to evaluate DMPs



# Four Stakeholders Resources & incentives

## Researcher

- **Resources to determine where data should go**
- Public rewards / recognition for data sharing
- Increased citations for future funding and job security

## Publisher

- Increased impact factors when data is reused
- Better tools for tracking data reuse & sharing
- **Resources to determine where data should go**

## Librarian

- Increased engagement with other stakeholders
- Better tools for tracking data reuse & sharing
- **Resources to determine where data should go**

## Funder

- FAIR education for reviewers & awardees
- Better tools for tracking data reuse & sharing
- Ability to track FAIR data & having quantifiable metrics



# AgBioData can bridge the workflow





# Questions for the community

- **In person & virtual breakout room participants**
  - If there was a tool or portal to help authors, publishers, and librarians determine where there data should go, what would it look like?



# Questions for the community

- **In person & virtual breakout room participants**
  - If there was a tool or portal to help authors, publishers, and librarians determine where their data should go, what would it look like?
  - Thinking of the barriers and incentives to making published data FAIR, are there things we have missed?



# Questions for the community

- **In person & virtual breakout room participants**

- If there was a tool or portal to help authors, publishers, and librarians determine where their data should go, what would it look like?
- Thinking of the barriers and incentives to making published data FAIR, are there things we have missed?
- **What are the most important challenges to address?**



# Thank you



Ruth Isaacson  
GSA



Leyla Cabugos  
Librarian, Cal Poly



Jenna Daenzer  
GSA



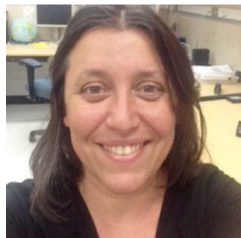
Leonore Reiser  
TAIR curator



Sook Jung  
Asst Research  
Professor. GDR



David Molik  
Computational  
Biologist,  
USDA ARS



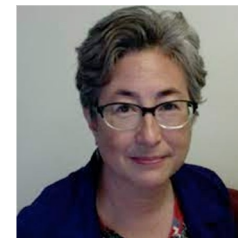
Daniela Raciti  
Exec Editor,  
microPublication &  
Wormbase Curator



Jacqueline Campbell  
Geneticist, USDA ARS  
SoyBase curator



Adam Wright  
Software Engineer  
Wormbase.Reactome



Karen Yook  
Exec Editor,  
microPublication &  
Wormbase Curator

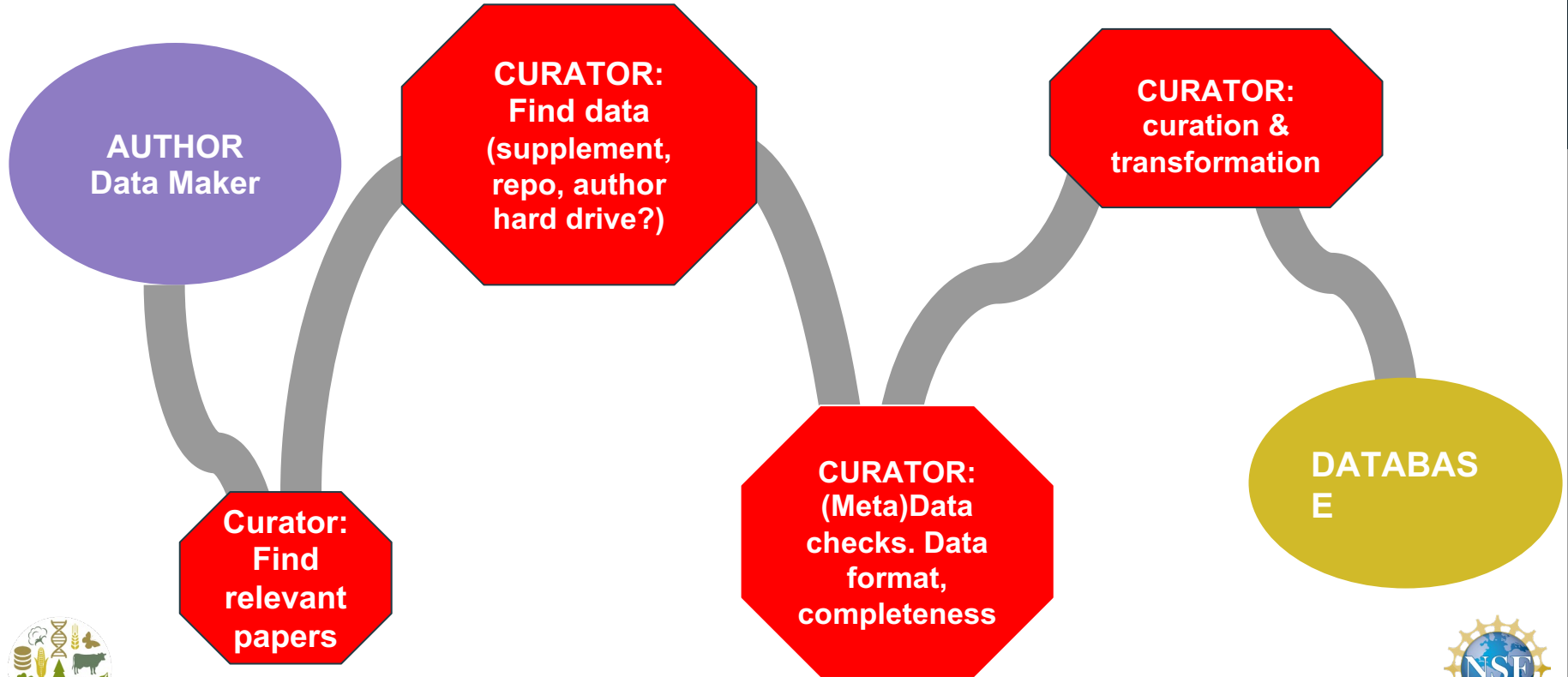


Daniel Morris , Professor (advisory)



# Supplemental Slides

# Current publication-curation workflow



# Example Stakeholder Focus: Genetic, Genomic and Breeding Database Curator

**Motivated By**

**Challenged By**

**Possible Incentives**

**Possible tools**



# Example Stakeholder Focus: Genetic, Genomic and Breeding Database Curator

**Motivated By**

**Challenged By**

**Possible Incentives**

**Possible tools**





# Stakeholder Focus: GGB Database Curator

## Motivated By

- Desire to present comprehensive, integrated data to the user community
- Need to maintain a desired resources that community values



# Stakeholder Focus: GGB Database Curator

## Challenged By

- Process data post-publication
- High volume of data & not enough curation time
- Data on request - lack of response from authors
- Poorly / Incorrectly formatted data and metadata



# Stakeholder Focus: GGB Database Curator

## Possible Incentives

- Increased community use
- Increased value by community
- Money for curation services



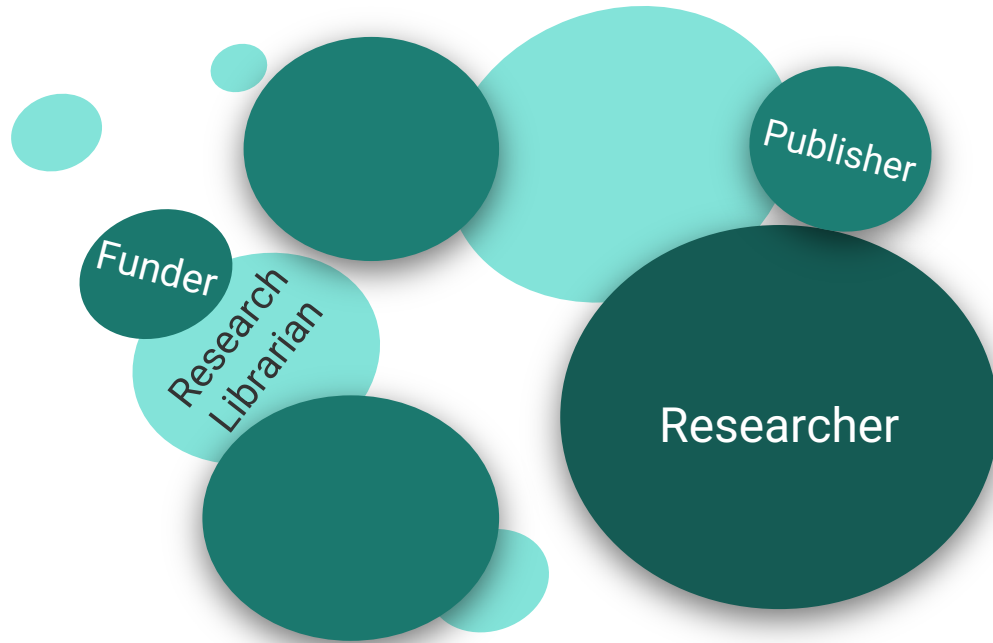
# Stakeholder Focus: GGB Database Curator

## Possible tools

- Software that facilitates data submission & a professional curator review as part of the publication pipeline
- Resource for authors & journals to determine where data should go
- Better metadata



Focus on 'before or during' publication not afterwards (FAIR from the start)



# FSLWG focus is on AgBioData community DBs

- Community databases host many different types of data from different sources
  - Common data types for a given species including those from other sources (e.g. sequences, RNA seq reads)
  - Unique datasets not found in other places (e.g G2P data, markers, mapping populations) or solely represented in supplemental data files.
- The value add for our user communities is
  - Streamlining the data finding and acquisition process
  - Curation and standardization
  - Integration with other data types



# Stakeholder Focus: Researcher

## Motivated By

- Professional rewards
- Recognition
- Moving science forward



# Stakeholder Focus: Researcher

## Challenged By

- Not knowing:
  - Where data should go
  - How to properly format data/metadata
- Time consuming
- Unclear benefits of data sharing





# Stakeholder Focus: Researcher

## Possible Incentives

- Public recognition for data sharing
- Increase citations for research



# Stakeholder Focus: Researcher

## Possible tools

- Software that facilitates data submission
- Better tools for tracking data reuse and reporting
- Accessible resources that authors can use to determine where data should go



# Stakeholder Focus: Funder

## Motivated By

- Advancing science through thoughtful allocation of funds.
- Increasing US national competitiveness and food security



# Stakeholder Focus: Funder

## Challenged By

- Not knowing all the places data should go
- Different program areas have different specialist knowledge
- Rely on reviewers who do not have the specific knowledge to evaluate DMPs



# Stakeholder Focus: Funder

## Possible Incentives

- Ability to track FAIR OS/OSTP memo compliance
- Having quantifiable metric to access value of funded research



# Stakeholder Focus: Funder

## Possible tools

- Better tools for tracking data reuse and reporting on how often data is shared
- Publicly accessible resources that authors and journals can use to determine where the data should go
- FAIR OS education for researchers

