AgBioData Ontology Working Group
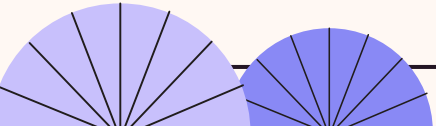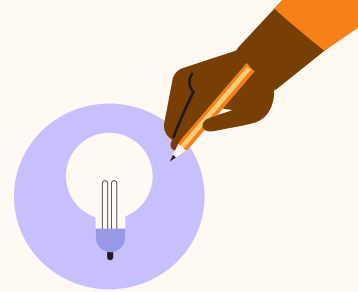
Laurel Cooper, Marie-Angélique Laporte
Co-chairs

2023 AgBioData Community Workshop

Overview

- Introduction to Ontologies
- The AgBioData Ontology WG
- WG Results
- Challenges

Simplified Generalization of Controlled Vocabulary Types

Ontology
Hierarchies
Relations
Axioms
Restrictions

Thesaurus
Hierarchies
Definitions
Relations

Taxonomy
Hierarchies

Glossary
Definitions

Precision / Semantic Strength / Complexity / Cost

The Ontology Working Group

- Started work in early Nov 2021

- Had biweekly meetings (~25) on Zoom through the end of Nov 2022

- Group size has varied from about 10 to 20

- Members spanned the globe- from New Zealand to North America to Europe and Great Britain

- Average meeting attendance was about 6.4 members

- A number of folks joined up, but never participated
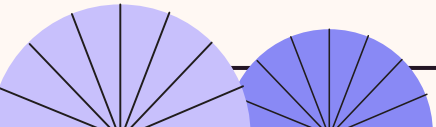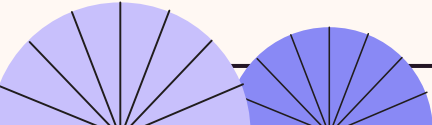
Core members
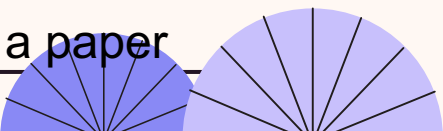
Erin Antognoli
Tanya Berardini
Amanda Cooksey
Damion Dooley
Peter Harrison
Sunita Kumari
Rex Nelson
Jonathan Sears
Taner Sen
Anne Thessen
Kathe Todd-Brown
Brandon Whitehead

Special thanks to Jennifer Woodward-Green who started out as chair

Ontology Working Group Objectives

- Survey the ontologies in use and compare to the results of previous survey, done in 2017 (along with Data Federation WG)

- Identify tools that are missing

- Identify barriers to ontology use

- Invite a well-known ontology expert for the webinar series

- Long-term goals were a set of recommendations and a paper

What did we accomplish?

- Survey on the use of ontologies by the community
- Paper on preparation with the Data Federation WG
- List of Use Cases for Ontologies in the AgBiodata Community, along with Identified Gaps and Barriers to Implementation
- Webinar on ontologies and FAIR data by Chris Mungall

Ontology survey results

- The goal of the survey was to better understand the ontologies used by the AgBiodata community: what are they and which ontologies are used to annotate what type of data
- Survey built on Google forms went out in July 2022
- Survey targeted to the AgBioData community
- Received responses from 33 individuals, representing 37 databases or resources
- Responses were skewed towards plant researchers (24/33)

A

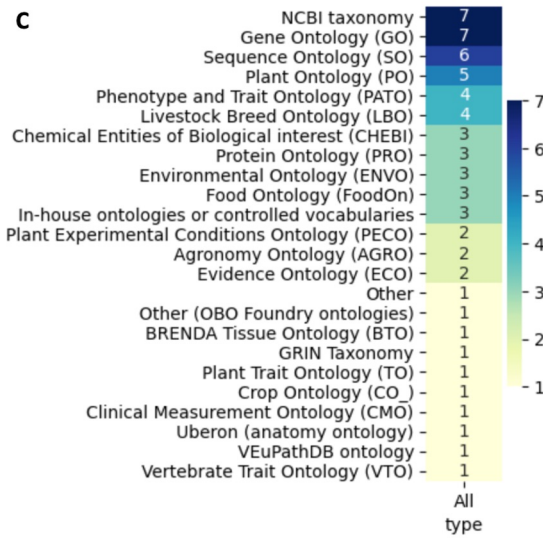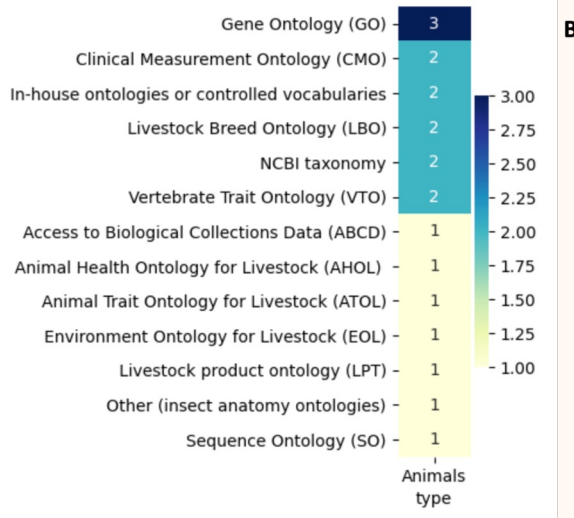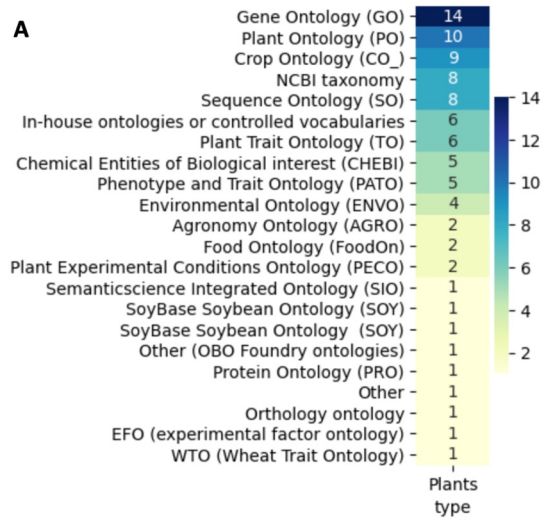| Ontology	Plants type
Gene Ontology (GO)	14
Plant Ontology (PO)	10
Crop Ontology (CO_)	9
NCBI taxonomy	8
Sequence Ontology (SO)	8
In-house ontologies or controlled vocabularies	6
Plant Trait Ontology (TO)	6
Chemical Entities of Biological interest (CHEBI)	5
Phenotype and Trait Ontology (PATO)	5
Environmental Ontology (ENVO)	4
Agronomy Ontology (AGRO)	2
Food Ontology (FoodOn)	2
Plant Experimental Conditions Ontology (PECO)	2
Semanticscience Integrated Ontology (SIO)	1
SoyBase Soybean Ontology (SOY)	1
SoyBase Soybean Ontology (SOY)	1
Other (OBO Foundry ontologies)	1
Protein Ontology (PRO)	1
Other	1
Orthology ontology	1
EFO (experimental factor ontology)	1
WTO (Wheat Trait Ontology)	1

B

Ontology	Animals type
Gene Ontology (GO)	3
Clinical Measurement Ontology (CMO)	2
In-house ontologies or controlled vocabularies	2
Livestock Breed Ontology (LBO)	2
NCBI taxonomy	2
Vertebrate Trait Ontology (VTO)	2
Access to Biological Collections Data (ABCD)	1
Animal Health Ontology for Livestock (AHOL)	1
Animal Trait Ontology for Livestock (ATOL)	1
Environment Ontology for Livestock (EOL)	1
Livestock product ontology (LPT)	1
Other (insect anatomy ontologies)	1
Sequence Ontology (SO)	1

C

Ontology	All type
NCBI taxonomy	7
Gene Ontology (GO)	7
Sequence Ontology (SO)	6
Plant Ontology (PO)	5
Phenotype and Trait Ontology (PATO)	4
Livestock Breed Ontology (LBO)	4
Chemical Entities of Biological interest (CHEBI)	3
Protein Ontology (PRO)	3
Environmental Ontology (ENVO)	3
Food Ontology (FoodOn)	3
In-house ontologies or controlled vocabularies	3
Plant Experimental Conditions Ontology (PECO)	2
Agronomy Ontology (AGRO)	2
Evidence Ontology (ECO)	2
Other	1
Other (OBO Foundry ontologies)	1
BRENDA Tissue Ontology (BTO)	1
GRIN Taxonomy	1
Plant Trait Ontology (TO)	1
Crop Ontology (CO_)	1
Clinical Measurement Ontology (CMO)	1
Uberon (anatomy ontology)	1
VEuPathDB ontology	1
Vertebrate Trait Ontology (VTO)	1

A. Ontologies that are used by the plant databases; B. Ontologies that are used by databases dealing with animal data. C. Ontologies that are used by databases storing both animal and plant data, as well as other types of data relevant to agriculture (e.g. fertilizers, …).

Ontology survey results

- Most respondents in the community are ontology users of externally managed ontologies and not creators
- Most respondents are contributing to ontologies, mainly through GitHub
- Ontologies are mainly used for annotation/curation and searching/archiving
- Main Barriers to the use of ontologies: insufficient funding and expertise to get started. Also mentioned, ontologies are insufficient or too specific

Ontology use cases
Understanding the gaps and barriers

Ontology Use Cases	Ontology Gaps	Ontology Use Barriers
Normalize conceptual knowledge in machine readable formatted metadata	Limited use of linked open data, URI	High complexity
Link conceptual knowledge across the WWW	Annotation time, skills	Ontology Curation and , Annotation time and skills
Annotation of experimental data	what ontologies are available/which to use?	Finding the appropriate tools/ontologies/databases/minimum reporting guidelines for experiments.
Ensuring data interoperability and supporting comparative analysis through common ontology use.	Managing community use of overlapping ontologies, need for mapping between ontologies.	More general ontologies often not appropriate for agriculture, ontology descriptions and terms not appropriate for agriculture.
Semantic linkage, linked open data (LOD)	Mappings, awareness of available tools (algorithms)	Lack of URI capacity in smaller vocabularies or groups
High throughput Phenotyping (HTP) annotation	New parameters being studied	

Ontology use cases
Understanding the gaps and barriers

Ontology Use Cases	Ontology Gaps	Ontology Use Barriers
Guidance for researchers (particularly those new to ontologies) in ontology use and appropriate tools		Lots of tools, lots of ontologies, very different ways to update ontologies and what that requires.
Expert review, engagement, contributions to vocabularies	Engagement of experts to contribute new concepts, definitions, or corrections	Many approaches by different groups. How to effectively source updates from community experts in an accessible form, automation, lower barriers to engagement.
Machine Learning, AI, and reasoning	Not all domains and concepts included	
Hypothesis generation		
Graph generation	Combining different ontologies	
Improving search	Needed ontology may not be available	A lot of work is needed to make useful in a specific database

Webinar

Chris Mungall

- The Gene Ontology: Making functional annotation of plants and animals FAIR.
- The Gene Ontology is one of the most widely used databases in the biosciences, covering functional annotation of genes and gene products across a wide range of species. The GO is ubiquitously used to analyse a variety of types of high-throughput experimental data. Originally created to unify functional annotation across a handful of model organism databases, the GO has grown to encompass more species, and the structure of the GO has been extended to integrate with other ontologies such as CHEBI and the Plant Ontology. The structure of annotations has also evolved, and the GO now includes more expressive pathway-oriented annotations in the form of GO-CAMs (Causal Activity Models). In this talk I will give a practical guide to the structure of GO, how to find and request terms, how to search and create annotations, and how to use GO tools. I will also talk about how the broader AgBioData can contribute to the GO consortium to help seed functional annotation efforts in a more diverse range of organisms, and in particular with agriculturally relevant species.

Challenges

- Attendance of the members
- Hard to find a use case that could be implemented, as WG members work on different topics and that would require a bigger time commitment

THANK YOU!

Do you have any questions?