FOUNDATION FOR FOOD & AGRICULTURE RESEARCH®

Summary of Responses to Request for Information: Data Solutions for Accelerating Climate-Smart Agriculture Research and Action

AgMission is a collaboration between the Foundation for Food & Agriculture Research and the World Farmers' Organisation to co-develop global, climate-smart solutions to achieve net zero emissions and bolster adaptation and resilience to climate change. Key to this effort is identifying barriers and closing knowledge gaps to accelerate adoption of farming and ranching practices that reduce and mitigate greenhouse gas emissions and promote resiliency. Over the past two years, AgMission has worked with consultants, experts and stakeholders to identify key areas where research investment can overcome these barriers. One emerging area of need is to enable agricultural research and decision support modeling to find, access and use the wealth of data and knowledge on agricultural systems and practices, including those adopted for purposes of climate change mitigation and adaptation. These data are heterogenous, lack standard metadata and are stored in a range of different public repositories or held privately; current efforts to transform this data into actionable knowledge remain fragmented.

In Fall of 2022, AgMission solicited responses to an open, public Request for Information (RFI) to further identify the status and needs for data standardization and integration in agricultural research to in support of climate-smart agriculture. Recognizing the interdisciplinary nature of climate-smart agriculture, and that integration and navigation solutions have the potential to provide broad advances and benefits, FFAR welcomed responses across the spectrum of agricultural sciences. We requested input on two major themes: Data Interoperability through Standardization and Data Synthesis through Navigation. The RFI was open from November 9 through December 21, 2022 and 17 sets of responses were received.

The majority of respondents were from Universities in the U.S., small to medium sized U.S.-based and international I/NGOs, and digital agricultural tool developers. In response to nine total questions, respondent comments provided a diverse and substantial set of resources, insights and recommendations. Respondents shared over 25 tools, websites and ongoing and developing initiatives. Especially abundant were referrals to resources related to emerging digital solutions, automation and FAIR data management. In addition, over 15 academic publications were shared with a particular focus on identifying existing ontologies and priority legacy datasets.

Summary of Responses: Data Interoperability Through Standardization

The purpose of this theme was to begin to identify the key necessary developments to unlock greater interoperability of agricultural data sets – including those collected in the course of research projects, model development and decision support applications. These include data collected by individual scientists, government agencies and private sector agronomy support services provided by companies, NGOs or other organizations. Given the wide diversity of data collection systems, standardization to the level that would enable machine-readable interoperability across data sets remains a challenge. One initial step would be to align data to common ontologies (data definitions) and metadata standards.

RFI respondents identified a number of specific and general categories for both further development of and alignment to existing metadata standards and ontologies. While ontologies have been developed for a wide range of subjects in agricultural research, a common theme in responses was the lack of awareness and adoption of existing ontologies by agricultural researchers, in large part because of a lack of training and resources to support the time required for learning and use. Respondents also noted that ontology libraries currently are overly complex and challenging for widespread use by researchers without specific training. Another important consideration was developing and communicating best practices for use of ontologies – for example, that ontologies need to be adopted early in a research project rather than waiting until data collection phases are completed.

Respondents had given substantial thought to where gaps exist in current ontologies as well as how to address the barriers to greater adoption of existing ontologies. Several respondents indicated ontologies that are consistent between disciplines, or coordination between different ontologies, are needed to capture the entire agriculture system (including soil, fertilizer characteristics, socioeconomic data, etc.). Alternatively, this need could be met by "data dictionaries" that provide translations across disciplines. Respondents indicated a need for systemic identification and filling of gaps through existing community forums (e.g., CGIAR) and a need for greater transparency in decisions regarding ontologies. The USDA Long-Term Agroecosystem Research (LTAR) network was identified as a success story for the adoption and application of ontologies to characterize data across diverse field locations and experiments.

"For most individuals to devote time to seeking out an appropriate ontology and learning to apply it to their work, the benefit would need to be apparent. Regrettably, the benefits are not immediately clear to many key groups that would be responsible for collecting data and using ontologies."

Specific ontology gaps identified:

- Organism occurrence (important for pest management and ecosystem services)
- Soil biology (taxonomy, microbial soil functions, DNA sequencing)
- Generic ontologies for crop production, supply chain and value-added traits
- Gaps in climate mitigation and adaptation research and other rapidly emerging domains, including novel crop traits
- Gaps in crop species, particularly specialty crops
- Animal waste management and manure characteristics
- Energy and water efficiency
- Controlled environment agriculture ontologies
- Landscape scale factors (beyond field/farm)

Solutions to Limited Ontology and Metadata Standard Adoption

A number of potential solutions to the identified barriers were discussed in the responses. Institutionally, integrating funding explicitly for data management and stewardship into research grants would facilitate the resources needed; funding institutions should provide additional guidance to grantees on available resources and clearly communicate expectations.

A number of innovations enabled by technology were also identified. First, to identify the appropriate ontologies, there are predictive analytic and machine learning techniques that can assist with automation. There are existing efforts to integrate such ontology discovery into data entry tools and therefore partially automate tagging of data as it is recorded. Several respondents indicated the potential for AI to develop ontologies, if needed, and to harmonize data post-collection. Finally, respondents the need was highlighted the need for an organized network or system of data repositories for scientists to easily identify where relevant data may be found.

Broadly speaking, researchers have higher awareness and adoption of metadata reference standards, terms and definitions than ontologies; however, RFI respondents also identified needs and barriers in this area. In particular, the lack of standardization and consistency in metadata use presents a barrier to interoperability of data collected. Respondents also identified a need to raise awareness of metadata standards in the research communities. A specific gap in metadata standards mentioned was the need for standards to credit traditional knowledge and provide biocultural context to agricultural data. Finally, one mechanism for addressing some of these challenges was the need for development of ing community awareness on the value of data re-use and the role of standards in making that research possible. Frequently, for example, data stored without meta-data cannot be sufficiently characterized to be used to parameterize agricultural models or in meta-analyses. Respondents also identified the lack of incentive and unclear value proposition to encourage researchers to undertake the time and effort to fully characterize their data with metadata.

A larger and more diverse set of technology tools to facilitate metadata appears to be available, particularly when looking broadly to other scientific disciplines for examples. While the potential to develop automatic or self-generative data annotation and ontology tools exists, these are not yet to the level needed, and development is occurring primarily in the private sector. Beyond the development of technology, a needs to be a cultural shift and adequate training and resourcing are needed so that scientists see the value in investing the time required to adopt metadata standards consistently.

“Data standards do not work well on their own. To succeed, they need more than one player and more than one tool. For data standards to take root, they need an environment that encourages collaboration, funding, infrastructure and policy development as well as people to champion them consistently.”

FOUNDATION FOR FOOD & AGRICULTURE RESEARCH®

Summary of Responses: Data Synthesis Through Navigation

One purpose behind supporting tools and efforts to standardize data for interoperability is to enable the use of advanced computational sciences to navigate across disparate datasets for data synthesis that can support advances in research as well as science-based solutions to agricultural challenges. These advances can enable the creation of web-based visual navigator tools to analyze structured and non-structured data and facilitate knowledge generation and analysis. Respondents were asked to reflect on the barriers and opportunities of such a navigational tool for advancing climate-smart agriculture research and adoption.

The barriers identified by respondents echoed some of the same barriers as identified in the previous sections; recurring themes were the lack of awareness and use of existing standards and ontologies, as well as the lack of training and resources to incentivize scientists to prioritize data stewardship. Repeatedly mentioned in responses was the need to demonstrate the value of ontologies, metadata and other data management practices that facilitate interoperability. When asked what the key requirements and characteristics were for data sets to support decision support tools on climate-smart agriculture, respondents indicated the importance of public data sets with high quality information from long-term experiments, soils data, including from chronosequence studies, and historical data of species occurrence. A specific recommendation was to convene a transdisciplinary working group with professional societies to articulate data needs for climate-smart agriculture. Respondents also identified some priority targets of historical or legacy datasets that could achieve greater use if data rescue efforts made them machine actionable.

Additional barriers specific to achieving interoperability were identified, including:

- Geospatial data methods are poorly matched to farm data so data sharing is unlikely to be in a standard format that facilitates interoperability
- FAIR data sharing can disincentivize collaboration with private citizens – the two are not typically viewed as compatible.
- Beyond available training, active mentoring of new staff on FAIR data is needed

Respondents generally indicated that while the cultural and institutional barriers to FAIR data remain, the technical challenges of interoperability remain secondary. A need for training programs and opportunities was mentioned repeatedly in responses, as was the need for long-term investment in infrastructure to ensure continuity over time.

Next Steps

Based on the responses received through this RFI, FFAR will be exploring several of the topics recommend that align with current research areas, in particular:

- Scoping a proof-of-concept project or set of projects to demonstrate the value of a) creating and adopting metadata standards and ontologies and b) utilizing interoperable data sets to investigate scientific questions and/or demonstrate climate solutions for farmers
- Explore the level of development of and potential need for automatic self-generative ontology and metadata tools for public sector research
- Evaluate the need for workforce development programs in data science for agricultural research
- Conduct follow up conversations to identify best practice guidance for FFAR grantees on the recommended and available standards, data repositories and other data resources they can use to meet FAIR data standards in their research projects
- Discussion of data stewardship and training will be incorporated into FFAR-wide discussions to determine potential relevance across all FFAR programs

If you are interested in submitting additional comments or feedback, please contact AgMission@foundationfar.org. Subscribe to AgMission news here to stay up to date on program developments.

Additional References and Resources Identified

- Meta-analysis of how ontologies are used, including what analyses they make possible and time savings in research
- Ontology look up service
- Juno Evidence Alliance
- Cultivating FAIR principles for agri-food data
- Enabling Open-source Data Networks in Public Agricultural Research
- Global agricultural concept space: lightweight semantics for pragmatic interoperability
- Sustainable solutions to end hunger

Resources and communities for data sciences for agricultural research

- CGIAR Ontology Community of Practice, AgroFIMS and GARDIAN
- USDA National Agricultural Library Ontology development
- ASABE Smart Farms linking research and industry for FAIR data
- Frontiers in Artificial Intelligence recently opened a research topic called "Semantics and Natural Language Processing in Agriculture"
- Example of AI used to create ontologies for crop nutrients
- Model specific data standard (DSSAT, AgMIP)
- OSU ontological approach for smart foodshed data

AI data applications in in other disciplines

- DivSeek Commons for Plant Genetic Resources
- ICICLE
- OSU data science

Examples of resources for automating metadata entry

- DataHarmonizer for pathogen genomics
- Biodiversity domain work
- OntoMaton to suggest ontology terms for data stored in Google sheets
- Texspreso for biological databases
- FairScribe and GEMS help users select ontology terms
- Tools to convert between metadata standards
- Publication on ontology-agnostic meta-data schema
- Report on data standards in soil