Towards designing an ontology encompassing the environment-agriculture-food-diet-health knowledge spectrum for food system sustainability and resilience.

Ruthie Musker, Matthew Lange, Allan Hollander, Patrick Huber, Nathaniel Springer, Courtney Riggle, James F. Quinn, Thomas P. Tomich.

University of California, Davis
Davis, California

Abstract— Feeding 9 billion people is not solely a matter of food, health, nutrition, and the environment. Promoting human health by increasing the sustainability and resilience of food systems requires integrating information from a broad range of disciplines from human nutrition/health systems and agricultural/natural systems to social, financial, physical and political systems. Ontologies serve to specify common terminologies for critical concepts and relationships within these systems, however very few ontologies have been developed with this interdisciplinary focus. Biological ontologies, whether focused on human physiology, soil quality, or nutritional value are only part of the story when it comes to determining linkages throughout the food system that help determine human health and well-being. We seek to build an ontology of food and food systems that encompasses the relevant sustainability issues in their entirety. We have already built an ontology of sustainable sourcing of agricultural raw materials issues and indicators, but aim to expand our ontology to include attributes of resilience, and other issues along the environment-agriculture-food-diet-health knowledge spectrum. Additionally, we aim to create this ontology with the intention of quick usability for the food system decision-maker.

Keywords— food systems ontology development, food system sustainability, food system resilience

I. Introduction

Ontologies are becoming increasing essential for synthetic research and statistical analytics where data from one scientific community is transformed through sequential models (a scientific workflow) to address problems typically posed by different communities accustomed to different terminologies for overlapping concepts. Important early efforts to formalize ontological frameworks include in biomedicine (NIH, OBO), agriculture (AGROVOC), and environmental biology (TDWG, NBII, GBIF), jointly characterized by rich vocabularies for names of chemical compounds, soils, species and varieties, and complex interactions. However, despite wide agreement that environmental quality is an important driver of sustainable food systems, which in turn promote health, the ontological efforts to date do not span the range of linked analyses needed to research the indirect but powerful connections that interact in these systems.

While increased understanding of the linkages between food, diet, sustainability and health offer solutions from individual to global health improvement, optimization, and maintenance; it would be shortsighted to decouple knowledge and information about food and the food system from the community that produces and utilizes it. Thus, our research is divided into two components. The first is to provide a practical, open access information platform that will bring coherence to a comprehensive array of food and health systems information. The second is to ensure that the information platform will sufficiently serve the environment-agriculture-food-diet-health system community it is meant to benefit.

By creating this ontology, we aim to help decision-makers access reliable data and metadata, observe justifications of relationships between concepts, and quickly see which other actors are interested in similar issues. The end goal of our ontology is three-fold:

1. Our ontology will provide decision-makers with **access to data discovery** and clearly show data that applies to their region of interest, what metadata exists, how to aggregate or disaggregate available data, how interoperable the data are based on data sources, etc.

2. Our ontology will focus on **conceptual linkages** among actors/concepts and the justifications/validations for those conceptual linkages. We believe that multiple linkages may exist between two issues due to various contexts, such as region, scale, commodity, and social structure.

3. Any framework covering this range of disciplines must have mechanisms to link to **standardized usages** (e.g., the Linked Data universe) and have mechanisms for cross-referencing identical or closely related concepts in standard use by different communities of practice.

These goals will in a format that is flexible based on the end users needs. This infrastructure, which supports standards-based loosely-coupled yet interoperable linked data will collapse time to cross-disciplinary insight and discovery, concomitantly accelerating food and health innovation.

Although recognition has been paid to creating an ontology of sustainability, the Sustainable Development Goals Interface Ontology (SDGIO), this ontology was built based on an already established conceptualization of sustainability goals, and rarely mentions "food" outside of "food security". Our approach builds from the inside-out, understanding that inclusivity of information, concepts, linkages, and data eliminates disciplinary boundaries and allows all food system stakeholders the ability to help create and use our ontology. Evidence shows that the "co-creation" of transdisciplinary research can improve credibility, relevance, and legitimacy, ultimately helping overcome social, administrative, and political boundaries and, thereby, improving chances for development of innovative strategies with better prospects for successful implementation [1][2].

II. Current Ontology

The ontology proposal builds upon the work of Springer et al. (2015) [3], and the conference paper of Musker et al. (2015) [4]. Foundational work at University of California, Davis by the Agricultural Sustainability Institute and the Information Center for the Environment screened international sustainability assessments, livelihoods assessments, and published food industry sustainability policies to identify a comprehensive set of food system sustainability issues and indicators, providing a starting point for environmental and food production portions of a food-and-health ontology. Springer et al. (2015) discuss the development of an ontology to express the relationship between environmental indicators, issues, conceptual frameworks, and data and reference sources (https://github.com/IC-FOODS/asi_sustainability/blob/master/ontology/sustsource.owl). This ontology contains classes for two conceptual frameworks, four sets of sustainability capital groups, and 44 integrated issues and describes relationships between indicators and issues using the SKOS schema.

Another approach to our current ontology takes a DPSIR framework approach [5] and determines structured linkage relationships between the Drivers, Pressures, States, Impacts and Responses (Figure 1). Each issue from Springer et al. (2015) could be placed in one of these categorizations, therefore creating a basic standardized ontology of the food system at the issue-to-issue level (Fig 1). Furthermore, we developed a structured typology of indicators to characterize the linkages between issues and indicators to determine where condition assessment, warning signal, threat identification, trend monitoring, or performance indicators would be most appropriate.

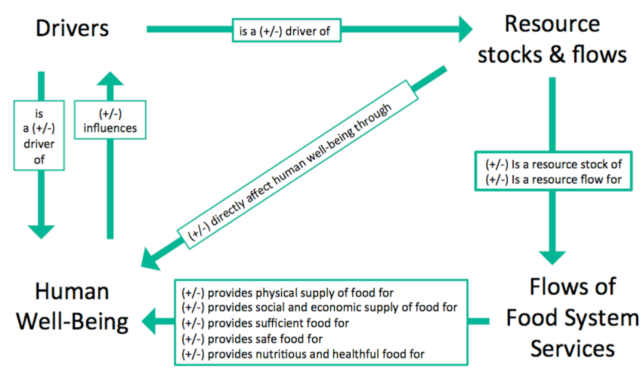

Fig 1. Issue-to-issue triples in a DPSIR framework.

Musker et al. (2015) incorporates the concept of resilience into Springer et al. 2015's sustainability ontology. The concept of resilience in food systems may assist with understanding both the current state of the food system and potential methods and strategies to increase food and nutrition security worldwide [6]. Our current conceptual ontology utilizes attributes of resilience [4] to discover strategies to alter the nature of a relationship or linkage through building resilience (Fig 2). For example: How can the relationship be strengthened? Reduced? Changed in direction? What capacity needs to be built to achieve the correct (or most sustainable) directionality of the relationship? How do you measure these capacities?

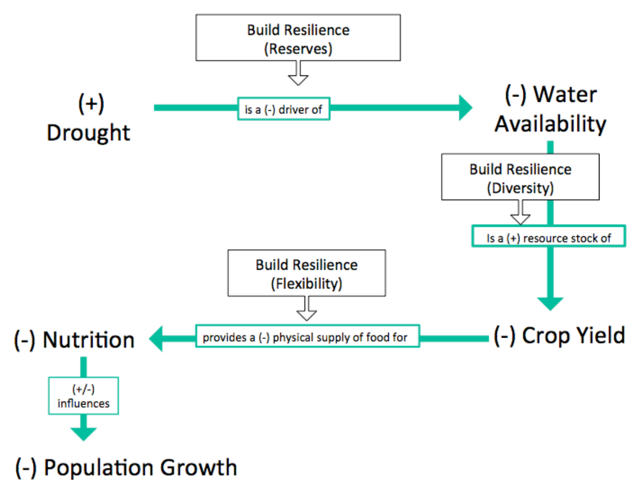

Fig 2. Building resilience as discovering strategies to change nature of relationships between issues.

We have described the Springer et al. (2015) and Musker et al. (2015) ontology in terms of RDF triples; however, our issues and concepts within our ontology can be expressed using OWL. The main concepts are Classes, Properties (subdivided into Object Properties and Data Properties) and Individuals. Springer et al. 2015 represented this information in the form of a web, however using Protégé, it is possible to translate this web of relationships into formal logical assertions. Since the component issues are expressed using terms from FAO's AGROVOC thesaurus, we have connected the URI of the term to our ontology. Currently, our ontology process is unconstrained, meaning any concept can be connected to any other concepts.

III. Next Steps in Ontology Development

Our food systems ontology will assist in classifying the communities of practice who will use the ontology, what kind of information they gather and types of information they seek. This information would be freely available to other actors, reducing unnecessary repetition in research, while assisting the creation of partnerships. Information, data, and results could be reviewed by others for decision-making processes to be transparent, and open access.

Our team is building upon the current ontology and including already existing ontologies, the Sustainable Development Goals Interface Ontology, Food On, Environment Ontology (ENVO), and National Center for Biomedical Ontology (NCBO), and will expand greatly to incorporate food, diets, eating, exercise and the human microbiome. This work will involve collaboration among academic and industry experts to ensure accuracy in our ontology development.

The boundaries of this ontology will focus on food and food systems, agricultural supply chains, ecosystem health, nutritional status of the global population, and the technology available to supply data about these knowledge domains.

IV. User-focused Ontology

While ontologies are indeed conceptually interesting; the power lies in the ontology's ability to facilitate information discovery and retrieval for the decision-maker. Due to this fact, it is crucial while building ontologies to consider the audience who requires the information, and could make the most use out of the ontology. We seek to understand the communication requirements necessary to allow the information discovered by ontology development to be useful quickly buy the target user group. Outlined are five ways that our ontologies will enable decision-makers to assess the information efficiently and effectively.

1. Communities of Practice

Our platform seeks to be useful to all communities of practice, within the environment-agriculture-food-diet-health knowledge spectrum. Communities of practice are defined as "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" [7]. We define these communities of practice as any stakeholder whose decisions positively or negatively affect the sustainability of the food system and include food manufacturing companies and corporations, academic researchers, consumers, policy makers, producers, etc. A successful platform may help the individual actor reconcile the different languages and conceptual assumptions of the multiple communities of practice within the food system in order to achieve something synthetic across the whole chain of causality.

2. Semantics

Currently, throughout the communities of practice, semantics create a real barrier to progress. Very different terminology is used for indicators, metrics, and fundamental issues. Development of semantics will simplify information and data discovery and can lead to indicator selection that is usable by all actors throughout agricultural supply chains and health systems. Our platform will greatly reduce this confusion by creating a library of health and food system terms, assessing where and how they are used and who uses them, that will allow the communities of practice to not only understand each other, but collect and share interoperable data. We will take advantage of existing vocabularies, such as CABT, AGROVOC, NALT, GEMET, and others to provide a foundation for semantic standards and development.

3. Completeness

Although communities of practice may use different language to describe the issues they care about, perceive different relationships among those issues, and use different indicators to measure them, there exists totality of sustainability issues that can be addressed by each community of practice.

However, the necessity of completeness could depend on the communities of practice and how they perceive boundaries. One option is, across these communities of practice, a linked global network of this sustainability information could act as a boundary of all the possible issues any given community could consider. Communities of practice can use such a global list to ensure that they are representing the complete list of issues for their cases. In this case, our ontology could demonstrate a large list of issues that must be taken into account in order to achieve a sustainability goal.

The second option takes the belief that completeness is not an achievable goal, and communities of practice exist in part because they are defined by conventions on what concepts/entities are intrinsic actors (state variable) and which are extrinsic drivers and there are no explicit boundaries on how far one could pursue a sequence of individually informative links. For this second option, a wide-ranging ontology like ours does the opposite – one can query on what kinds of nodes and edges are (say) one or two steps from those included in an analysis or decision tree, look at the list, and ask what important processes/entities have been excluded and how much one cares about those sources of incompleteness.

At the same time, such a comprehensive list can be unwieldy to sort through, especially with key differences depending on semantics, scale, scope, sector, commodity focus, etc. One way of minimizing the amount of information needed to represent all key issues while simultaneously assuring completeness is to apply an optimization algorithm to the information set defined by the stakeholder group to solve the "minimum covering set" (MCS) problem. Such an algorithm selects the minimum set of indicators from a broader set of known indicators required to represent the set of issues that are deemed important by the a particular community. Different algorithms can be used to solve the MCS problem, including a heuristic minimization approach [8] and an integer programming method [9]. Regardless, communities of practice can use such a tool to consistently ensure completeness and consideration of tradeoffs while selecting indicators to represent the issues they care about.

4. Workflows and processes

Workflows are essential when utilizing data and information that is traceable, transferable and interoperable among many communities of practice. Such a process has been successful in the conservation community at creating "best-practice" decision-support systems for conservation projects [10]. Information technology tools such as MIRADI (https://miradi.org/) help conservation partners develop boundaries, measurements, goals, and strategies for specific uses that can be shared across user groups without a loss of generality. If a certain process works well for a certain communities, this process can be documented, maintained, and utilized by other communities as well.

New knowledge generated through scientific research, cultural exchange, social development, and practical experience is constantly revealing new issues as other issues are addressed. Consistent approaches to identify sustainability issues of importance and indicators to measure them must be flexible enough to incorporate emerging issues, insights, and data sources [3]. We continue to explore how explore directionality among concepts and applications of scalability within, or complementary to, our ontology.

V. Use Case

Currently, the Agricultural Sustainability Institute at UC Davis is working with the Sacramento Area Council of Governments to determine conservation land use planning potential. A wide range of stakeholders are involved, including local government planners, landowners/developers, federal and state resources regulators, environmental advocacy organizations, and agricultural groups/individual landowners. Many of the stakeholders have similar concerns over specific issues and desires for interoperable data but do not understand each other's needs or have a place to express these concerns. The goal of this use case is to test our ontology of sustainability to provide guidance on the collection of existing data and modeling needed to underpin an effort to define regional strategies for the conservation of natural resources and support for management of those lands for environmental and economic benefits.

VI. Conclusion

Our ontology will be hosted at the University of California, Davis, at the new International Center Food Ontologies, Operability, Data, and Semantics (IC-FOODS), and will have support from the Agricultural Sustainability Institute, Food Science and Technology, the Information Center for the Environment, and the Innovation Institute for Food and Health. Plans are underway to host the information facility in the university digital library, in collaboration with the UC Davis Data Science Initiative. We are hoping to bring several communities of practices into our ontology development, both in the ontology creation side, and the user side. These communities of practice include regional and national governments, academics, industry professionals, non-profit organizations, and foundations, with expertise from data security and curation to domain expertise in human health and nutrition, agricultural supply chains, and intellectual property.

Questions we will be able to help communities of practice answer:

1. What widely-measured variables documenting human, social, and environmental health are interconnected in a food system and nutritional context? What processes, actors, and measures/indicators definitively need to be included in a conceptual framework for understanding impacts of food system on health and well-being, and where does one set defensible system boundaries?

2. How can we classify and formalize terminology for these connections to help relevant actors to exchange and integrate information efficiently and accurately, to make well-informed decisions regarding economic, social, and environmental dimensions of sustainability? What data does one need, and where are the data gaps?

3. How can data be most effectively curated to maintain the integrity of rigorous scientific

research surrounding human nutrition and health, together with economic, social and environmental sustainability, while ensuring that these data are interoperability among disciplines, actors, sectors, and scales?

Upon completion of our ontology, these are examples of practical questions that users would be able to address:

1. How to quantify sustainability of a food product (e.g., a chocolate bar) from a social, human nutrition, environmental and economic perspective?

2. Which indicators could describe the current state of sustainability for, as an example, Switzerland's imported goods? Which indicators could be used to identify potential shocks to Switzerland's food system and ways to monitor these vulnerabilities?

3. How can existing data be used to understand the link between food prices, human health outcomes, and droughts?

4. Which network of actors in the food system has the ability to make positive changes in the supply chain of a commodity in a specific production region, (e.g.: dairy milk in North America)?

REFERENCES

[1] E. Ostrom, Burger, J., Field, C.B., Norgaard, R.B., Policansky, D., 1999. Revisiting the commons: local lessons, global challenges. Science 284, 278–282. doi:10.1126/science.284.5412.278

[2] W. C. Clark, Tomich, T.P., Noordwijk, M. van, Guston, D., Catacutan, D., Dickson, N.M., McNie, E., (2011). Boundary work for sustainable development: natural resource management at the Consultative Group on International Agricultural Research (CGIAR). Proc. Natl. Acad. Sci. 200900231. doi:10.1073/pnas.0900231108

[3] N. P. Springer, Garbach K, Guillozet K, Haden VR, Hedao P, Hollander AD, et al. (2015) Sustainable Sourcing of Global Agricultural Raw Materials: Assessing Gaps in Key Impact and Vulnerability Issues and Indicators. PLoS ONE 10(6): e0128752. doi:10.1371/journal.pone.0128752

[4] R. Musker, Springer N, Quinn J, Hollander A, Huber, P, Riggle C, Tendall D, Grant M, Six J, Tomich T. (2015). Applying ontologies to sustainability and resilience concepts in food systems. Conference Paper: Global Food Security: Adaptation, Resilience, and Risk. University College London. September 10, 2016.

[5] Ash, N., Blanco, H., & Brown, C. (2010). *Ecosystems and human well-being: a manual for assessment practitioners*. Island Press.

[6] D. M Tendall, J. Joerin, B. Kopainsky, P. Edwards, A. Shreck, Q.B. Le, P. Kruetli, M. Grant, and J. Six. (2015). "Food System Resilience: Defining the Concept." *Global Food Security* 6 (October): 17–23. doi:10.1016/j.gfs.2015.08.001.

[7] E. Wenger. 2011. "Communities of Practice: A Brief Introduction," October. http://scholarsbank.uoregon.edu/xmlui/handle/1794/11736.

[8] P. R. Huber, Springer, N. P., Hollander, A. D., Haden, V. R., Brodt, S., Tomich, T. P., & Quinn, J. F. (2015). Indicators of global sustainable sourcing as a set covering problem: an integrated approach to sustainability.*Ecosystem Health and Sustainability*, *1*(2), art7.

[9] Balinski, M.L. and R E Quandt. (1964). On an integer program for a delivery problem. Oper. Res. 12, 300–304. doi:10.1287/opre.12.2.300

[10] The Conservation Measures Partnership. (2013). Open standards for the practice of conservation (No. Version 3.0).