Towards a Global Food Systems Datahub

Editorial

Hande Küçük McGinty [a,*], Cogan Shimizu [b], Pascal Hitzler [a] and Ajay Sharda [c]
[a] *Department of Computer Science, Kansas State University, KS, USA*
E-mails: hande@ksu.edu, hitzler@ksu.edu
[b] *Department of Computer Science and Engineering, Wright State University, OH, USA*
E-mail: cogan.shimizu@wright.edu
[c] *Department of Biological and Agricultural Engineering, Kansas State University, KS, USA*
E-mail: asharda@ksu.edu

Keywords: Global Food Systems, Knowledge Graphs

In the quest for agricultural sustainability, we face the challenge of feeding the global population under the constraints of finite resources and a delicate ecological balance. The intricate interplay of climate dynamics, socio-economic factors, and environmental stewardship demands an approach to agriculture that is as intelligent and adaptive as it is respectful of our planet's capacities. Central to this endeavor is the synthesis and utilization of vast, heterogeneous datasets that span from crop genomics to market trends, and from soil health to consumer preferences. Yet, the current paradigm is fragmented, with valuable data isolated across domains, lacking the coherence and accessibility needed for actionable insights. While there is an urgent imperative: to enable the kind of data-driven decision-making that can foster sustainability in agriculture, traditional methods of data analysis are insufficient. They lack the depth and agility required to navigate the complexities of modern agriculture.

For example, precision agriculture technologies have rapidly evolved over the last two decades. Almost all of the machine systems including planters, air- seeders, liquid application systems, and crop harvesters have extended stacks of sensing and control systems to automatically control the application of crop inputs [1]. Farmers can potentially access numerous databases which can be classified into soils (e.g., spatial soil features); machine application (e.g., crop input flow from sensors, and geodata); climate (e.g., temperature, precipitation); remote sensing (e.g., satellite, manned, and unmanned aerial systems with sensors); machine vision (e.g., autonomous scout vehicles, and cameras on large equipment) and above- or below- ground sensors providing near real-time data (e.g., soil moisture, temperature) [1, 2].

However, the unprecedented amount of data availability has not radically transformed data-centric technologies like variable rate technologies (seeding, nutrition), site-specific application (chemical selection, pest control, plant nutrition). Among other factors including noisy data, incomplete datasets, missing data layers, and errors in data, the lack of data interoperability has been the biggest obstacle preventing data integration and analytics for decision support [7]. The data sources have been developed by various manufacturers using non-standardized file formats, units, and metadata; have varying level of both spatial and temporal resolutions; need specific softwares to read different data file; and no common platform for data curation, integration, spatial and temporal normalization, and analytics. The magnitude of data gathering, integration, and analytics further magnified due to the transformation of rural landscape, land ownership, gradual aging (the average age of a farmer in the U.S. is 57.5 years), and continuous decline in human capital remaining in rural America available for crop production [5]. Furthermore, while humans may seamlessly transition between different levels of spatial, or temporal, or informational granularity, knowledge representation in AI is by its nature rigid and traditionally forces a priori defined perspectives on the data model.

*Corresponding author. E-mail: hande@ksu.edu.

1570-0844/$35.00 © 2024 – IOS Press. All rights reserved.

Consequently, knowledge representation is usually used to solve a singular and specific problem or across a set of very similar purposes. This produces significant overhead when preparing data for subsequent processing or analysis, as the resulting knowledge bases cannot be readily reused. At large scale, and in contexts such as climate-adaptive AI, where relevant data is highly heterogeneous, the cost of data preparation and management, and of assurance of data quality is often prohibitive.

To enable climate-adaptive agriculture, it is necessary to establish modular and multi-granular knowledge representation solutions at large scale that are agnostic towards different levels of spatial information and temporal granularities [3, 4, 6]. Therefore, the necessity for a coherent and standardized approach to agricultural data is not just a technological imperative, but also a socioeconomic one, as it holds the potential to usher in a new era of efficiency and sustainability in agriculture. One transformative solution would be a Global Food Systems Datahub powered by a knowledge graph backend which would aspire to transcend the limitations listed above by establishing a robust framework for integrating and interpreting data at a granular level. This frame would be beyond an assembly of information; but a living, evolving digital ecosystem that maps relationships and infers patterns, thereby enabling advanced AI applications in the realm of digital agriculture since knowledge representation is a long-standing AI discipline that is currently gaining large-scale adoption in industry in the form of KGs [5].

The need for a knowledge graph-powered hub is multi-faceted. At its core, the knowledge graph offers a semantically rich structure that mimics the interconnectedness of real-world systems, allowing for a nuanced representation of data relationships. This facilitates the application of foundational AI research areas such as multi-modal spatiotemporal deep learning, explainable AI, and neuro-symbolic question answering. With these technologies, we can analyze and understand the function and malfunctions of agricultural systems, predict outcomes, and make recommendations that are transparent and justifiable.

In practical terms, the Global Food Systems Datahub stands to revolutionize how we approach sustainable intensification – the enhancement of production on existing farmland in harmony with the environment. AI-enabled tools developed from the hub's knowledge graph can inform strategies that balance crop yield with ecological conservation, such as the spatial targeting of prairie strips or the integration of agrivoltaic systems. High-resolution data can inform landowners' decisions about land use, optimizing for both production and ecosystem service provision. Moreover, this hub would bridge the gap between the data-rich insights of individual landowners and the broader needs of program managers and policymakers, who often operate with a paucity of detail. It will render visible the hidden layers of agricultural data, thus enabling targeted interventions and informed decision-making at a micro-scale, with macro-scale implications.

The overarching need for the datahub aligns with the pressing global need of sustainable development. The global food systems datahub can play a pivotal role in mitigating climate impacts, managing water-energy-food nexus challenges, and supporting socio-economic and cultural adaptability in agricultural practices. In this scope, for example, falls a recently started large internally funded project at Kansas State University titled "Towards a Global Food Systems Data Hub: Seeding the Center for Sustainable Wheat Production" (Hitzler and McGinty involved) which is set to seed the beginnings of such a data hub by initially focusing on sustainable wheat production. In a similar vein, the National Science Foundation's Proto-OKN program[1] is working on a network of interconnected knowledge graphs, some of which are set to bear significant importance to a future global food systems datahub. Many other related efforts are of course under way world-wide.

In this special issue, several KG centric approaches have been presented by researchers that are also relevant to the overall goal.

- *Dimitris Zeginis, Evangelos Kalampokis, Raul Palma, Rob Atkinson, Konstantinos Tarabanis, A Semantic Meta-Model for Data Integration and Exploitation in Precision Agriculture and Livestock Farming* present their study that aims to identify the characteristics of data used in precision agriculture and livestock farming and the user requirements related to data modeling and processing from nine real cases at the agriculture, livestock farming and aquaculture domains. They propose a semantic meta-model that is based on W3C standards (DCAT, PROV-O and QB vocabulary) in order to enable the definition of metadata that facilitate the discovery, exploration, integration and accessing of data in the domain.

[1] https://www.proto-okn.net/

- *Christopher Brewster, Nikos Kalatzis, Barry Nouwt, Han Kruiger, Jack Verhoosel, Data Sharing in Agricultural Supply Chains: Using semantics to enable sustainable food systems* propose a data sharing architecture, the Ploutos, that is built on three principles a) reuse of existing semantic standards; b) integration with legacy systems; and c) a distributed architecture where stakeholders control access to their own data. Their system has been developed based on the requirements of commercial users and is designed to allow queries across a federated network of agrifood stakeholders. The Ploutos semantic model is built on an integration of existing ontologies and is built on a discovery directory and interoperability enablers, which use graph query patterns to traverse the network and collect the requisite data to be shared. Their data sharing approach is highly extensible with considerable potential for capturing sustainability related data.
- *Damion Dooley, Liliana Andrés-Hernández, Georgeta Bordea, Leigh Carmody, Duccio Cavalieri, Lauren Chan, Pol Castellano-Escuder, Carl Lachat, Fleur Mougin, Francesco Vitali, Chen Yang, Magalie Weber, Hande Kucuk McGinty, Matthew Lange, OBO Foundry Food Ontology Interconnectivity* focus on reuse of existing efforts and standardized vocabulary in modeling food and agricultural data. Interoperability features stem from food-related ontologies that are part of, or align with, the comprehensive Open Biological and Biomedical Ontology (OBO) Foundry platform. This study shows how these features can be leveraged by research organizations and businesses for their projects or data exchange initiatives. The demand for a standardized vocabulary spans various aspects of the food supply chain, including agricultural production, harvesting, preparation, processing, marketing, distribution, and consumption. This standardization extends to the broader impacts on health, economics, food security, and sustainability, requiring consistent analysis and reporting tools. This study emphasizes and shows that to meet the need for a controlled vocabulary, it is crucial to develop domain-specific ontologies. The development of these ontologies involves close coordination among curators to establish recommended patterns for food system vocabulary.
- *Damion Dooley, Magalie Weber, Liliana Ibanescu, Matthew Lange ,Lauren Chan, Larisa Soldatova, Chen Yang, Robert Warren, Cogan Shimizu, Hande Kucuk McGinty, William Hsiao, Food Process Ontology Requirements* show that some of this standardized vocabulary efforts should focus on food processing related efforts. Their work focuses on the ontology components – object and data properties and annotations – needed to model food processes or more general process modelling within the context of the Open Biological and Biomedical Ontology (OBO) Foundry and congruent ontologies. They show that these components can be brought together in a general process ontology that can be specialized not only for the food domain but for carrying out other protocols as well. Many operations involved in food identification, preparation, transportation and storage – shaking, boiling, mixing, freezing, labeling, shipping – are actually common to activities from manufacturing and laboratory work to local or home food preparation, which can help better model global food systems data.
- *Katherine Thornton, Kenneth Seals-Nutt, Mika Matsuzaki, Damion Dooley, Reuse of the FoodOn Ontology in a Knowledge Base of Food Composition Data* focused on reusing FoodOn ontology with their specialized food composition database, WikiFCD. WikiFCD is a structured knowledge base dedicated to the composition and characteristics of various food items. Their primary objective was to adopt FoodOn identifiers for consistent referencing of food items within WikiFCD. They show that the benefits of this integration extend beyond internal improvements. For the broader FoodOn community, the enriched data from WikiFCD—which provides detailed composition information at the food item level—offers substantial potential advantages. Notably, WikiFCD is linked to Wikidata and includes a SPARQL endpoint capable of supporting federated queries. This connection enables researchers to perform comprehensive and sophisticated queries that span across WikiFCD and Wikidata platforms, leveraging cross-domain data. Such queries significantly broaden the scope of possible data analyses and insights, particularly in the realms of nutritional science, food safety, and supply chain management. This integration not only streamlines data handling within WikiFCD but also enriches the FoodOn community's resources, making it a mutually beneficial advancement in food data management and application.

With these papers in the special issue, we see that the community is already recognizing the benefits of standardized vocabulary generation, semantic modeling of global food systems data and they may be used for further analyses and reuse by industry and researchers. We hope that this special issue highlights the global food system

as one of the most essential yet complex constructs, underpinning not just health and nutrition but also socioeconomic structures, environmental balance, and cultural identities. It embodies an extensive range of data types, from climatological observations, agricultural practices, supply chain logistics, to consumer behavior and nutritional information.

This multidisciplinary combination, however, poses a significant challenge: the integration and interoperability of diverse datasets that remain siloed across different domains. The existing gap hinders the potential for multi-layered analysis, comprehensive understanding, and efficient decision-making, which are critical in addressing pressing issues like food security, sustainability, and resilience against climate change. Advancements in artificial intelligence (AI) and machine learning (ML) offer unprecedented opportunities to bridge this gap by enabling the translation of massive, heterogeneous data into a format that is not only comprehensible across various disciplines but also operable by computational systems. Therefore, the need for a Global Food Systems Datahub (GFSD) with a knowledge graph backend is our envisioned solution as a transformative platform that encapsulates the wealth of global food data into an interconnected, semantically-rich, and AI-ready infrastructure.

Acknowledgement McGinty, Hitzler and Shimizu acknowledge partial support by the National Science Foundation under award 2333532, *Proto-OKN Theme 3: An Education Gateway for the Proto-OKN*.

References

[1] S. Badua, A. Sharda, R. Strasser and I. Ciampitti, Ground speed and planter downforce influence on corn seed spacing and depth, *Precision Agriculture* **22** (2021), 1154–1170.

[2] J. Fabula, A. Sharda, J.D. Luck and E. Brokesh, Nozzle pressure uniformity and expected droplet size of a pulse width modulation (PWM) spray technology, *Computers and Electronics in Agriculture* **190** (2021), 106388.

[3] P. Hitzler and C. Shimizu, Modular Ontologies as a Bridge Between Human Conceptualization and Data, in: *Graph-Based Representation and Reasoning – 23rd International Conference on Conceptual Structures, ICCS 2018, Edinburgh, UK, June 20-22, 2018, Proceedings*, P. Chapman, D. Endres and N. Pernelle, eds, Lecture Notes in Computer Science, Vol. 10872, Springer, 2018, pp. 3–6. doi:10.1007/978-3-319-91379-7_1.

[4] H.K. McGinty, KNowledge Acquisition and Representation Methodology (KNARM) and Its Applications, PhD thesis, University of Miami, 2018.

[5] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson and J. Taylor, Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done, *Queue* **17**(2) (2019), 48–75.

[6] C. Shimizu, K. Hammar and P. Hitzler, Modular ontology modeling, *Semantic Web* **14**(3) (2023), 459–489.

[7] E.L. White, J.A. Thomasson, B. Auvermann, N.R. Kitchen, L.S. Pierson, D. Porter, C. Baillie, H. Hamann, G. Hoogenboom, T. Janzen et al., Report from the conference,'identifying obstacles to applying big data in agriculture', *Precision Agriculture* **22** (2021), 306–315.