Advancing Agriculture through Semantic Data Management

Editorial

Pascal Hitzler, [a] Krzysztof Janowicz, [c] Ajay Sharda, [b] Cogan Shimizu [a]
[a] *Department of Computer Science, Kansas State University, USA*
[b] *Department of Biological and Agricultural Engineering, Kansas State University, USA*
[c] *Geography Department, University of California, Santa Barbara*

Fundamentally, agriculture is about sustainably cultivating the environment to meet societal needs. However, neither the environment nor society are static or uniform. Instead, they vary across regions and time, and they form complex interaction networks. For instance, changing cultural norms may require an adjustment of practices even though these may not strictly be optimal from an agronomic perspective. Conversely, society has to adapt to changes in the environment, e.g., to ensure the long-term sustainability of natural resources. Decision-makers also need to account for regional aspects and interactions between neighboring regions that, to date, are often considered in isolation.

For example, the Ogallala Aquifer [1] is a part of the U.S. High Plains Aquifer System, spans eight states of the Great Plains, and provides water for a third of all irrigated land in the United States, while also supplying drinking water for millions of Americans. Despite various initiatives, the aquifer is still depleting as reductions in water usage due to precision agriculture are offset by new demands, such as biofuel and increasing environmental stress. While the Ogallala Aquifer is unique in its role for the U.S., it is prototypical for the complex intertwined relationships across the biotic, abiotic, and cultural factors that characterize agriculture like no other domain. While the aquifer's water levels are rising in Nebraska, they are declining in Kansas, New Mexico, and parts of Texas. A changing climate will further exaggerate these regional differences. The usage of water also differs among states ranging from serving the irrigation needs of rural America and the drinking water needs of urban America. Even water use rights differ among the states, e.g., granting Texans unrestricted rights to the water beneath their properties.

In the past, such conflicting interests and a societal consensus around topics such as environmental sustainability, tail docking, or genetically engineered foods have been addressed via commissions, elections, and regulations to reach joint explanations of new norms. Increasingly, decision-making in agriculture is too rapid, too multivariate, and too interlinked to be satisfactorily settled in such ways. Instead, more and more decisions are left to machine learning models and their supporting sensor networks that provide a wide range of heterogeneous data at multiple scales. However, current artificial intelligence models and precision agriculture techniques alone cannot readily capture the breadth of conflicting actors, interests, environmental factors, and regional differences while improving climate adaptation and sustainable intensification. And most importantly, they cannot provide explanations.

The discussion just provided makes it apparent that modern and sustainable agricultural decision making needs to be based not only on multi-faceted and multi-sourced, and thus highly heterogeneous data, but also needs to be supported by artificially intelligent decision support systems that can flexibly adapt to contextual factors based on knowledge about situational parameters, their relevance, and their implications.

To further illustrate this point, consider U.S. agriculture, which is a flourishing and robust industry con-

1570-0844/21/$35.00

tributing US$390 billion per year in annual revenue from agricultural commodities [2]. The top 10 commodities contributing 77% of this revenue among others include corn, soybean, wheat, chickens, cattle, and hay. Most of these crops are grown over very large acres with varied climate, soil, irrigation water, soil nutrition, pests, extent of technology, and level of intelligence used in crop production decision making. As an example, corn and soybean alone captures 41% of total cultivated farmland (1.5 million km^2), with an annual operating cost of US$48 billion [3, 4]. Over the last two decades, precision agriculture technologies have been systematically integrated for crop production, with current machines being bigger, wider and faster. These developments in agriculture, improved genetics, and enhancements in technology design have helped to increase farm productivity and yields. However, today's grand challenge as highlighted by the United States Department of Agriculture is to increase food production by 40% while cutting the environmental footprint by 40%.

Total farmland in the U.S. has steadily decreased from 3.8 million km^2 in 2000 to 3.6 million km^2 in 2019 [5]. In order to increase food production from limited farmlands, radical changes in decision making based on integrated digital data needs to be utilized to take every plant to its optimal yield potential. One of the key impediments to accomplish this task has been the gaps in site-specific decision making. Decision making for agricultural ecosystems to drive decisions has been becoming increasingly complex since it utilizes diverse data layers including soil, topography, water, crop, machine, pest, disease, and changing environment. However, these vast spatial and temporal digital data layers have not yet been utilized to develop AI decision making algorithms, because data layers are lacking integrability, spatial and temporal density, completeness, accuracy, accessibility, and availability due to privacy.

Comprehensively addressing agricultural needs such as those described above can be achieved by refinement and application of a broad range of Semantic Web technologies. We discuss some of the main pillars.

Semantic Data Integration As we have seen above, to address modern agricultural needs it is necessary to integrate large-scale, multi-sourced data from (sometimes sporadic) data streams in order to make this integrated data available for analysis. The Semantic Web field has provided research and solutions for this for decades [6], but they need to be tailored to the specifics of agriculture, and they need to scale both in terms of data size and speed. Complex temporal and spatial aspects play a major role, both of which are topics that have so far not received sufficient attention in research and solutions around ontologies, linked data, and knowledge graphs.

Semantic Data Enrichment Large volumes of relevant data, such as air quality, weather, or land use data, are already available, and sometimes even in the form of knowledge graphs. Additional large volumes of data are or will soon be created by agricultural sensor networks and autonomous agricultural machinery. In order to make use of this data, it needs to be annotated with sufficient semantic metadata to facilitate automated data integration and analysis at the required speed and in different and possibly changing environments of data streams. The same piece of data-producing equipment will be used in many different agricultural and data contexts, meaning different requirements on content, precision and resolution of the streamed data. We need to work towards an understanding of the exact requirements in each context, and towards conceptually and technologically scalable and sustainable solutions on how to meet different metadata requirements cost-efficiently in different scenarios and at scale.

Semantic Sustainable Data Management Data solutions will have to be in place that can be utilized long-term, and this requires emphasis on aspects that appear to be underrepresented in Semantic Web research. What are good and scalable solutions to evolve an ontology (as knowledge graph schema) while maintaining access and usability of legacy data [7, 8]? How to make decisions which data to keep long-term and in what format? How to develop data integration solutions that easily adapt to data, sensor and requirements contexts that change and evolve over time? Can our current ways of knowledge engineering cope with effects of semantic aging?

Knowledge-adaptive Data Analytics Collecting and integrating relevant data is a central aspect, as outlined above. However, in order to utilize this data, analytics capabilities need to be able to make use of a context in a flexible way. This includes, ideally, geographic and environmental factors, as well as socio-cultural factors such as local preferences, guidelines, and policies, and some of these may change more or less rapidly over time. Data analytics, currently dominantly reliant on machine learning methods, is at this time ill-equipped to make significant use of relevant and changing background context, and more research efforts are required on this front. From a Semantic Web context, a lead question is how to make systematic use of semantically rich and evolving metadata, for machine learning and analytics.

Semantic Explainability [9, 10] Furthermore, analytics solutions will have to be trusted by farmers, who may query system recommendations, in particular if they may not align with past experience or practice. Explanations of data analytics results will have to be provided in terms understandable by laypersons, which means that they have to be at a suitable level of abstraction from the raw data. While explainability, in particular in the context of machine learning, is being researched, the nature of the explanations is often in very basic terms, e.g. by highlighting parts of the input data that contributed most to the system's output. In these cases, it is still left to the human user to make sense of this. It would be much more helpful to have explanations expressed in terms that have more direct and immediate meaning within a particular domain.

The arguments just laid out provide us with some guidelines as to where the Semantic Web field needs to evolve to address the agricultural – and other similarly complex – challenges. It is necessary to develop solutions that are fit for long-term complex and changing settings, and that seamlessly interface with data analytics. Much of the current Semantic Web research, in contrast, is driven by short-term projects and individual capabilities, disregarding the additional complexities introduced by a complex application setting such as agriculture.

Acknowledgements. Authors Hitzler, Janowicz and Shimizu acknowledge support by the National Science Foundation under Grant 2033521 A1: KnowWhereGraph: Enriching and Linking Cross-Domain Knowledge Graphs using Spatially-Explicit AI Technologies. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Chaudhuri and S. Ale, Trends in groundwater contamination and salinization in the Ogallala aquifer in Texas, *Journal of Hydrology* **513** (2014), 376–390, doi:10.1016/j.jhydrol.2014.03.033.

[2] U.S. Department of Agriculture – Economic Research Service (2021). Annual cash receipts by commodity, Accessed: 2021-06-03. https://data.ers.usda.gov/reports.aspx?ID=17832.

[3] U.S. Department of Agriculture – Economic Research Service (2021). Commodity Costs and Returns, Accessed: 2021-06-03. https://www.ers.usda.gov/data-products/commodity-costs-and-returns/commodity-costs-and-returns/#Historical%20Costs%20and%20Returns:%20Hogs.

[4] U.S. Department of Agriculture – Economic Research Service (2021). Acreage, Accessed: 2021-06-03. https://www.nass.usda.gov/Publications/Todays_Reports/reports/acrg0620.pdf.

[5] U.S. Department of Agriculture – National Agricultural Statistical Services (2020). Farms and Land in Farms, 2019 Summary, Accessed: 2021-06-03. https://www.nass.usda.gov/Publications/Todays_Reports/reports/fnlo0220.pdf.

[6] P. Hitzler, A Review of the Semantic Web Field, *Commun. ACM* **64**(2) (2021), 76–83. doi:10.1145/3397512.

[7] C. Shimizu, K. Hammar and P. Hitzler, Modular Ontology Modeling, Technical Report, 2021, under Review. https://daselab.cs.ksu.edu/publications/modular-ontology-modeling.

[8] E. Blomqvist, K. Hammar and V. Presutti, Engineering Ontologies with Patterns – The eXtreme Design Methodology, in: *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi and V. Presutti, eds, Studies on the Semantic Web, Vol. 25, IOS Press, 2016, pp. 23–50. doi:10.3233/978-1-61499-676-7-23.

[9] P. Hitzler, F. Bianchi, M. Ebrahimi and M.K. Sarker, Neural-symbolic integration and the Semantic Web, *Semantic Web* **11**(1) (2020), 3–11. doi:10.3233/SW-190368.

[10] I. Tiddi, F. Lécué and P. Hitzler (eds), *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, Studies on the Semantic Web, Vol. 47, IOS Press, 2020. ISBN 978-1-64368-080-4.