

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 Amer-

ica. 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-

1 tributing US\$390 billion per year in annual revenue
2 from agricultural commodities [2]. The top 10 com-
3 modities contributing 77% of this revenue among oth-
4 ers include corn, soybean, wheat, chickens, cattle, and
5 hay. Most of these crops are grown over very large
6 acres with varied climate, soil, irrigation water, soil nu-
7 trition, pests, extent of technology, and level of intel-
8 ligence used in crop production decision making. As
9 an example, corn and soybean alone captures 41% of
10 total cultivated farmland (1.5 million km²), with an
11 annual operating cost of US\$48 billion [3, 4]. Over
12 the last two decades, precision agriculture technolo-
13 gies have been systematically integrated for crop pro-
14 duction, with current machines being bigger, wider and
15 faster. These developments in agriculture, improved
16 genetics, and enhancements in technology design have
17 helped to increase farm productivity and yields. How-
18 ever, today's grand challenge as highlighted by the
19 United States Department of Agriculture is to increase
20 food production by 40% while cutting the environmen-
21 tal footprint by 40%.

22 Total farmland in the U.S. has steadily decreased from
23 3.8 million km² in 2000 to 3.6 million km² in 2019 [5].
24 In order to increase food production from limited farm-
25 lands, radical changes in decision making based on in-
26 tegrated digital data needs to be utilized to take every
27 plant to its optimal yield potential. One of the key im-
28 pediments to accomplish this task has been the gaps
29 in site-specific decision making. Decision making for
30 agricultural ecosystems to drive decisions has been be-
31 coming increasingly complex since it utilizes diverse
32 data layers including soil, topography, water, crop, ma-
33 chine, pest, disease, and changing environment. How-
34 ever, these vast spatial and temporal digital data lay-
35 ers have not yet been utilized to develop AI decision
36 making algorithms, because data layers are lacking in-
37 tegrability, spatial and temporal density, completeness,
38 accuracy, accessibility, and availability due to privacy.

39 Comprehensively addressing agricultural needs such
40 as those described above can be achieved by refine-
41 ment and application of a broad range of Semantic
42 Web technologies. We discuss some of the main pil-
43 lars.

44 **Semantic Data Integration** As we have seen above,
45 to address modern agricultural needs it is necessary to
46 integrate large-scale, multi-sourced data from (some-
47 times sporadic) data streams in order to make this in-
48 tegrated data available for analysis. The Semantic Web
49 field has provided research and solutions for this for
50
51

1 decades [6], but they need to be tailored to the specifics
2 of agriculture, and they need to scale both in terms
3 of data size and speed. Complex temporal and spa-
4 tial aspects play a major role, both of which are top-
5 ics that have so far not received sufficient attention in
6 research and solutions around ontologies, linked data,
7 and knowledge graphs.

8 **Semantic Data Enrichment** Large volumes of rel-
9 evant data, such as air quality, weather, or land use
10 data, are already available, and sometimes even in the
11 form of knowledge graphs. Additional large volumes
12 of data are or will soon be created by agricultural
13 sensor networks and autonomous agricultural machin-
14 ery. In order to make use of this data, it needs to be
15 annotated with sufficient semantic metadata to facili-
16 tate automated data integration and analysis at the re-
17 quired speed and in different and possibly changing
18 environments of data streams. The same piece of data-
19 producing equipment will be used in many different
20 agricultural and data contexts, meaning different re-
21 quirements on content, precision and resolution of the
22 streamed data. We need to work towards an under-
23 standing of the exact requirements in each context, and
24 towards conceptually and technologically scalable and
25 sustainable solutions on how to meet different meta-
26 data requirements cost-efficiently in different scenar-
27 ios and at scale.

28 **Semantic Sustainable Data Management** Data solu-
29 tions will have to be in place that can be utilized long-
30 term, and this requires emphasis on aspects that ap-
31 pear to be underrepresented in Semantic Web research.
32 What are good and scalable solutions to evolve an on-
33 tology (as knowledge graph schema) while maintain-
34 ing access and usability of legacy data [7, 8]? How
35 to make decisions which data to keep long-term and
36 in what format? How to develop data integration so-
37 lutions that easily adapt to data, sensor and require-
38 ments contexts that change and evolve over time? Can
39 our current ways of knowledge engineering cope with
40 effects of semantic aging?
41

42 **Knowledge-adaptive Data Analytics** Collecting and
43 integrating relevant data is a central aspect, as outlined
44 above. However, in order to utilize this data, analytics
45 capabilities need to be able to make use of a context in
46 a flexible way. This includes, ideally, geographic and
47 environmental factors, as well as socio-cultural factors
48 such as local preferences, guidelines, and policies, and
49 some of these may change more or less rapidly over
50 time. Data analytics, currently dominantly reliant on
51

1 machine learning methods, is at this time ill-equipped
2 to make significant use of relevant and changing back-
3 ground context, and more research efforts are required
4 on this front. From a Semantic Web context, a lead
5 question is how to make systematic use of semantically
6 rich and evolving metadata, for machine learning and
7 analytics.

8 **Semantic Explainability** [9, 10] Furthermore, ana-
9 lytics solutions will have to be trusted by farmers,
10 who may query system recommendations, in particu-
11 lar if they may not align with past experience or prac-
12 tice. Explanations of data analytics results will have
13 to be provided in terms understandable by laypersons,
14 which means that they have to be at a suitable level of
15 abstraction from the raw data. While explainability, in
16 particular in the context of machine learning, is being
17 researched, the nature of the explanations is often in
18 very basic terms, e.g. by highlighting parts of the in-
19 put data that contributed most to the system's output.
20 In these cases, it is still left to the human user to make
21 sense of this. It would be much more helpful to have
22 explanations expressed in terms that have more direct
23 and immediate meaning within a particular domain.

24 The arguments just laid out provide us with some
25 guidelines as to where the Semantic Web field needs to
26 evolve to address the agricultural – and other similarly
27 complex – challenges. It is necessary to develop solu-
28 tions that are fit for long-term complex and changing
29 settings, and that seamlessly interface with data ana-
30 lytics. Much of the current Semantic Web research, in
31 contrast, is driven by short-term projects and individ-
32 ual capabilities, disregarding the additional complexi-
33 ties introduced by a complex application setting such
34 as agriculture.
35

36
37 *Acknowledgements.* Authors Hitzler, Janowicz and
38 Shimizu acknowledge support by the National Science
39 Foundation under Grant 2033521 A1: KnowWhere-
40 Graph: Enriching and Linking Cross-Domain Knowl-
41

1 edge Graphs using Spatially-Explicit AI Technologies.
2 Any opinions, findings, and conclusions or recommen-
3 dations expressed in this material are those of the au-
4 thors and do not necessarily reflect the views of the
5 National Science Foundation.
6

7 8 References

- 9
10 [1] S. Chaudhuri and S. Ale, Trends in groundwater con-
11 tamination and salinization in the Ogallala aquifer in
12 Texas, *Journal of Hydrology* **513** (2014), 376–390,
13 doi:10.1016/j.jhydrol.2014.03.033.
14 [2] U.S. Department of Agriculture – Economic Research Service
15 (2021). Annual cash receipts by commodity, Accessed: 2021-
16 06-03. <https://data.ers.usda.gov/reports.aspx?ID=17832>.
17 [3] U.S. Department of Agriculture – Economic Research
18 Service (2021). Commodity Costs and Returns, Ac-
19 cessed: 2021-06-03. [https://www.ers.usda.gov/data-products/
20 commodity-costs-and-returns/commodity-costs-and-returns/
21 #Historical%20Costs%20and%20Returns:%20Hogs](https://www.ers.usda.gov/data-products/commodity-costs-and-returns/commodity-costs-and-returns/#Historical%20Costs%20and%20Returns:%20Hogs).
22 [4] U.S. Department of Agriculture – Economic Research Service
23 (2021). Acreage, Accessed: 2021-06-03. [https://www.nass.
24 usda.gov/Publications/Todays_Reports/reports/acrg0620.pdf](https://www.nass.usda.gov/Publications/Todays_Reports/reports/acrg0620.pdf).
25 [5] U.S. Department of Agriculture – National Agricultural Sta-
26 tistical Services (2020). Farms and Land in Farms, 2019
27 Summary, Accessed: 2021-06-03. [https://www.nass.usda.gov/
28 Publications/Todays_Reports/reports/fnl0220.pdf](https://www.nass.usda.gov/Publications/Todays_Reports/reports/fnl0220.pdf).
29 [6] P. Hitzler, A Review of the Semantic Web Field, *Commun.*
30 *ACM* **64**(2) (2021), 76–83. doi:10.1145/3397512.
31 [7] C. Shimizu, K. Hammar and P. Hitzler, Modular Ontology
32 Modeling, Technical Report, 2021, under Review. [https://
33 daselab.cs.ksu.edu/publications/modular-ontology-modeling](https://daselab.cs.ksu.edu/publications/modular-ontology-modeling).
34 [8] E. Blomqvist, K. Hammar and V. Presutti, Engineering On-
35 tologies with Patterns – The eXtreme Design Methodology, in:
36 *Ontology Engineering with Ontology Design Patterns – Foun-
37 dations and Applications*, P. Hitzler, A. Gangemi, K. Janow-
38 icz, A. Krisnadhi and V. Presutti, eds, Studies on the Semantic
39 Web, Vol. 25, IOS Press, 2016, pp. 23–50. doi:10.3233/978-1-
40 61499-676-7-23.
41 [9] P. Hitzler, F. Bianchi, M. Ebrahimi and M.K. Sarker, Neural-
42 symbolic integration and the Semantic Web, *Semantic Web*
43 **11**(1) (2020), 3–11. doi:10.3233/SW-190368.
44 [10] I. Tiddi, F. Lécué and P. Hitzler (eds), *Knowledge Graphs for*
45 *eXplainable Artificial Intelligence: Foundations, Applications*
46 *and Challenges*, Studies on the Semantic Web, Vol. 47, IOS
47 Press, 2020. ISBN 978-1-64368-080-4.
48
49
50
51