Case Study Review

Introduction

This paper discusses the agriculture ontology created by Malik et al. (2015). The methodology and outcome are both critiqued and found to be lacking in rigor and completeness. The ontology is, however, a useful foundation that can be extended and adapted to provide solutions to real-world problems. Two examples are suggested; optimising the agriculture supply chain and building relationships between agriculture and food health to enable farming for health.

Ontology development approaches

Malik et al. (2015) propose creating a *domain ontology* for the agriculture domain based upon AGROVOC (Food and Agriculture Organization of the United Nations, 2023a). The motivations include no *generic ontology* exists for the agriculture domain and ontologies overcome the limitations of keyword searches through semantic relationships. Bimba et al. (2016) say domain ontologies align terminology and concepts so information can be shared, whilst generic ontologies are more useful across multiple domains, so although domain and generic ontologies were used interchangeably within the paper, it was a domain ontology that was developed.

The seven steps to develop an ontology (Noy & McGuinness, 2001) are quoted, but then not robustly implemented. For example, step 1 is “determine the domain and scope of the ontology” with a key question being “**for what we are going to use the ontology?**” (Noy & McGuinness, 2001: 5), however, the motivation appears to be little more than “one doesn’t already exist”. Key applications are referred to later in the paper, however, they remain generic and it seems like Malik et al. (2015) were looking for ways to use the ontology they had already proposed rather than defining a problem first.

Malik et al. (2015: 739) assert that AGROVOC “does not consider relationships between related entities”. Perhaps that was true in 2015, but today AGROVOC does hold relationships. For example, Brassica oleracea var. capitata (or cabbage) has a “has pathogen” relationship with five pathogens (figure 1). The relationships are not exhaustive though, missing important relationships such as fertiliser to soil and plant type, so the challenge with using AGROVOV without an ontology remains accurate.

... > Brassicales > Brassicaceae > Brassica > Brassica oleracea > Brassica oleracea var. capitata

PREFERRED TERM	Brassica oleracea var. capitata
BROADER CONCEPT	Brassica oleracea (en)
ENTRY TERMS	*Brassica oleracea capitata (en)* *cabbage (plant) (en)*
HISTORY NOTE	Updated preferred label from Brassica oleracea capitata to Brassica oleracea var. capitata in 2022. (en)
HAS PATHOGEN	Erysiphe cruciferarum (en) Erysiphe polygoni (en) Pyrenopeziza brassicae (en) Sclerotinia minor (en) Sclerotinia sclerotiorum (en)
HAS TAXONOMIC RANK	variety (taxa) (en)
PRODUCES	cabbages (en) red cabbages (en)
EDITORIAL NOTE	Author: L. (en)

Figure 1. AGROVOC except for Brassica oleracea var. capitata

(Food and Agriculture Organization of the United Nations, 2023b)

Malik et al. (2015) refer to “agriculture” throughout, which includes livestock (National Geographic, 2023), but they build an ontology excluding livestock with no mention as to why. They also mention consulting domain experts, but then go straight into defining the classes using “common terms that common people think of” (Malik et al., 2015: 740), finally concluding that as they continue to build the ontology they can “seek expert advice wherever necessary” (Malik et al., 2015: 741). This suggests that the initial ontology was built using AGROVOC as a data source, showing adherence to step 2 (reuse) from Noy & McGuinness (2001), but without any domain expertise to understand the nuances of the data. Kendal & Green (2019) provide a useful approach to knowledge acquisition, which postdates Malik et al. (2015), however, the principles would still have been well known. To improve the ontology they could have started with a series of unstructured interviews with domain experts to overview the agriculture domain before moving into more structured interviews using sample data from AGROVOC to fully understand the domain before jumping into defining classes, subclasses, and instances.

Methodology weaknesses aside, the ontology that was developed for the subset of agriculture excluding livestock has merit. It has taken the AGROVOC data set and created 5 classes; plant, disease, pest, pesticide, and fertiliser, along with subclasses and instances, and relationships between them.

Application areas and business context

The benefits stated by Malik et al. (2015) were superficial. Here, specific uses of the agriculture ontology are examined within their business context.

The first example is optimising the agriculture supply chain. Denis et al. (2020) explain that agriculture supply chains are complex, with multiple steps from farms to silos to transformation plants to other silos and then on to clients, with many decisions at each step (figure 2). Hundreds of thousands of permutations are impacted by internal factors such as farms and crops, and external factors such as weather and demand-driven pricing. They make a case for using digital twins to model the physical supply chain to simulate decisions and events to optimise the physical supply chain.

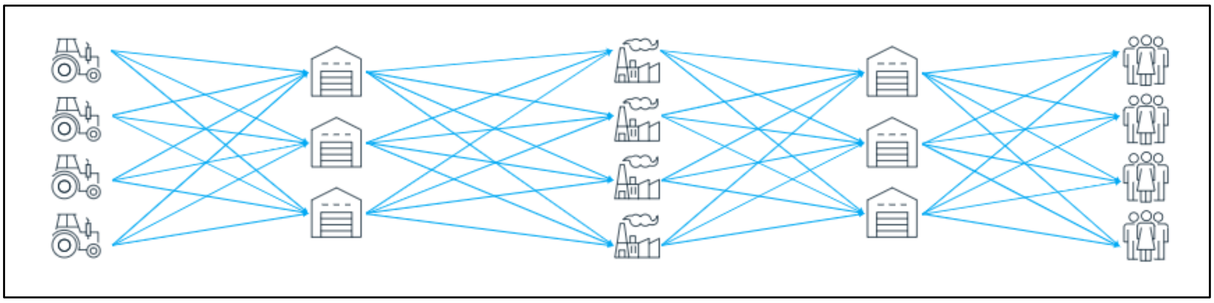

Figure 2. Multiple decisions in agriculture supply chains (Denis et al., 2020)

Singh et al. (2021) propose using ontologies to structure data before creating digital twins. They demonstrated improvements in creating and maintaining digital twins because the relationships are maintained within the ontology.

Pylianidis et al. (2021) found that the use of digital twins in agriculture is still relatively uncommon, but with potential. They suggest that operations can be automated as well as simulated by taking data from sensors, which would seek to automate some of the decisions noted by Denis et al. (2020).

Similarly, Ngo et al. (2018) propose using the abundance of sensor data from Internet of Things (IoT) devices being increasingly used in agriculture to create an IoT subdomain to sit alongside subdomains for agriculture, geography, and business to create a complete agriculture ontology (figure 3).

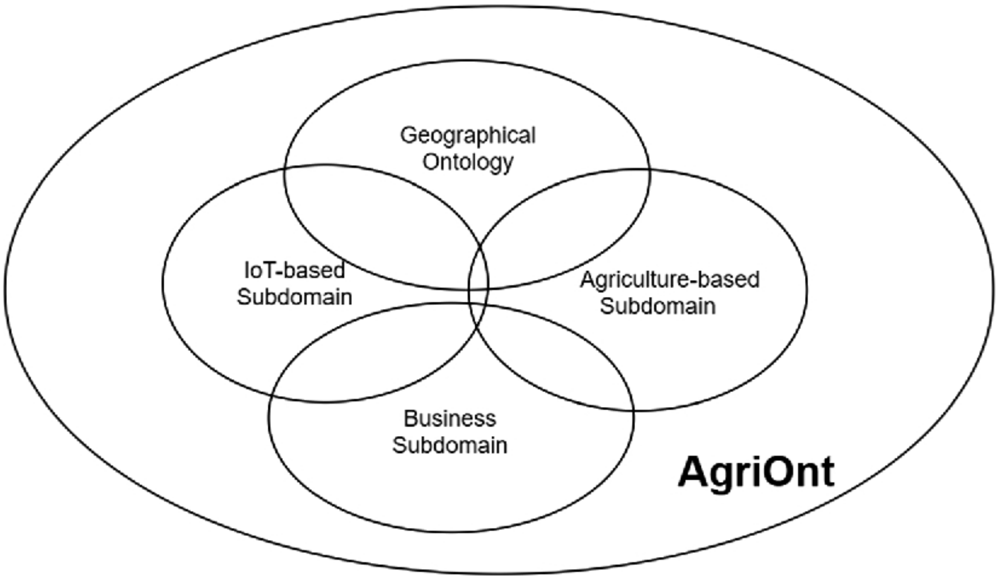

Figure 3. Agriculture ontology (Ngo et al. 2018: 179)

Looking at those opportunities together, the ontology devised by Malik et al. (2015) could be the agriculture subdomain, which when combined with the IoT, geographical and business subdomains would create the overall agriculture ontology that would enable the modelling and automation of agriculture supply chains, including the creation of digital twins. It would enable all parts of the supply chain to model, plan and optimise their decisions, such as the most appropriate pesticide and best time to spray, the optimal sequence to harvest different crops, and optimal silo utilisation. It would re-plan in response to external factors such as changing weather conditions. By continuously optimising the entire supply chain, yield would be maximised, and transport and storage costs would be minimised, driving a significant increase in profitability across the supply chain.

Another application of the agriculture ontology is to pursue a healthier relationship with food. Lange et al. (2007: 1432-1433) assert that “obtaining knowledge about the health in food” is difficult because “agriculture and health speak different languages”. They propose a schematic of ontologies to relate agriculture to food, nutrition, and health (figure 4).

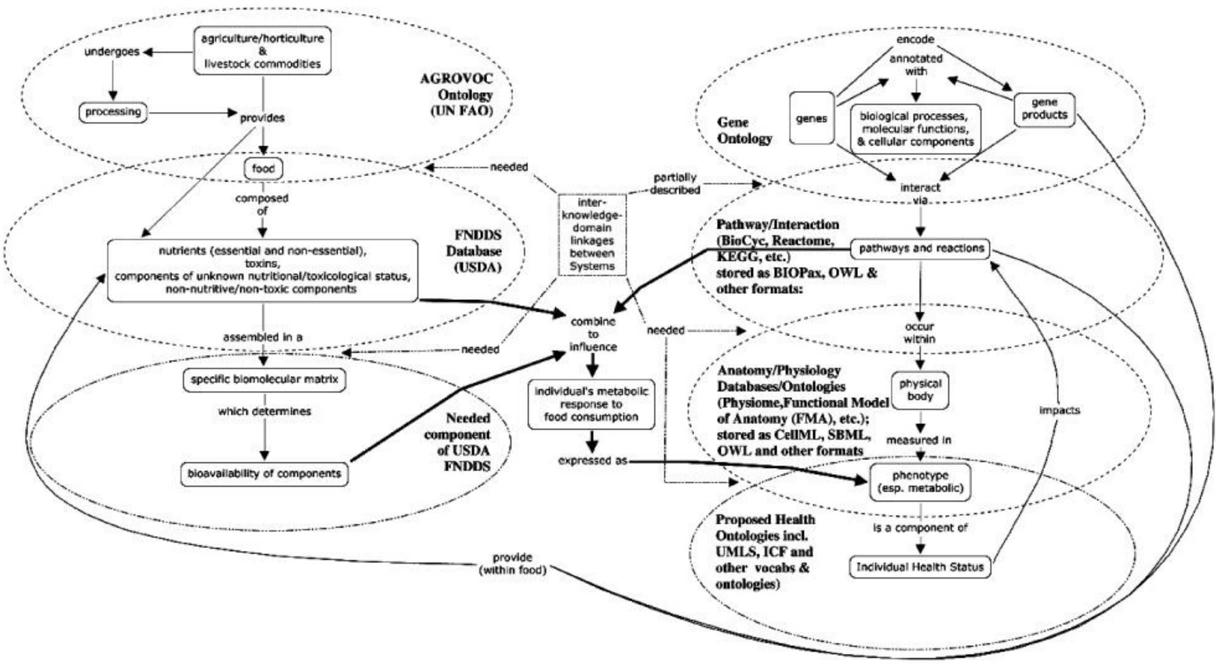

Figure 4. Schematic of ontologies for a Foods-for-Health knowledge system (Lange et al. 2007: 1432)

The schematic starts with an agriculture ontology based upon AGVROVOC and combines it with other ontologies to create a knowledge base that can map ingredients to food, the bioavailability of nutrients, and ultimately health.

> “Logical connectivity between the life-science knowledge domains will not only allow for richer and more meaningful exchanges and querying of data across the disciplines, but will also ultimately lead to the development of systems capable of automated reasoning about dietary interventions

> and directed health outcomes at the level of the individual.” (Lange et al. 2007: 1433).

Whilst Lange et al. (2007) are concerned with knowledge to inform consumer choice, it could go further. By understanding the implications of ingredients on health at a macro-scale, interventions in agriculture could be taken to promote healthier living. By understanding what is used in healthy foods, including choices of pesticides, alongside the availability of such ingredients and the popularity of the food, all of which can change over time, the agriculture industry could provide the right amount of the right produce to promote healthy living at an optimal cost. That could be through self-regulation or tax incentives at the agriculture level rather than tax at the point of sale.

Adapting to the application areas

Improvements to the ontology development methodology such as formal engagement of domain experts have already been discussed, so here the focus is on further adapting the agriculture ontology to maximise usefulness to the identified applications.

The supply chain example combined four ontologies, including agriculture (figure 2). However, certain data are missing from Malik et al (2015)’s ontology that would limit the relationships with the geographic ontology. Livestock should be added as already mentioned. Furthermore, additional relationships not adequately included in AGCROVOC should be added to expand the knowledge base and make the relationships with the geographic ontology more useful.

Chatterjee et al. (2019) propose a scheme for taking agriculture terms and unstructured data from websites and using natural language processing (NLP) and regular expressions to automatically find and extract relationships grows_in(Crop, Soil), found_in(Soil, Region) and suffers_from(Crop, Disease). They achieved precision and recall of over 80% apart from found_in which had precision of 61.7% and recall of 87.48%. This demonstrates that the agriculture ontology could be expanded automatically and continuously, by including carefully selected online sources.

The agriculture ontology from Malik et al. (2015) is similarly deficient for the healthy eating schematic. AGROVOC contains limited information about food, for example, is_made_from(bacon, pork) is there, but packaged foods are not. Nor are processing methods such as pork being cured to make bacon. The ontology would need to be extended to include this additional information, and since it is not all in AGROVOC there would need to be consultation with domain experts. Once complete, the additional relationships would allow the agriculture ontology to relate to the Food and Nutrient Database for Dietary Studies (FNDDS), or any other food and nutrient database.

Conclusions

The agriculture ontology (Malik et al., 2015) is deficient in both method and content. It is, however, a useful initial ontology based upon AGROVOC.

The ontology should be improved by extending the breadth to include livestock and the depth to include more relationships. Some data are available in AGROVOC, but many are not so domain experts should be consulted through interviews and potentially data could be supplemented with NLP on carefully selected online sources.

The improved ontology could be used as a base, on which other ontologies and capabilities could be built, to address genuine challenges in agriculture such as supply chain management and farming for healthy eating.

Word count: 1,650

References

Bimba, A.T., Idris, N., Al-Hunaiyyan, A., Mahmud, R.B., Abdelaziz, A., Khan, S. & Chang, V. (2016) Towards knowledge modeling and manipulation technologies: A survey. *International Journal of Information Management* 36(6): 857–871. https://doi.org/10.1016/j.ijinfomgt.2016.05.022

Chatterjee, N., Kaushik, N. & Bansal, B. (2019) Inter-subdomain relation extraction for agriculture domain. *IETE Technical Review* 36(2): 157-163. https://doi.org/10.1080/02564602.2018.1435312

Denis, N., Dilda, V., Kalouche, R. & Sabah, R. (2020) Agriculture supply-chain optimization and value creation Available from: https://www.mckinsey.com/industries/agriculture/our-insights/agriculture-supply-chain-optimization-and-value-creation#/ [Accessed 17 September 2023].

Food and Agriculture Organization of the United Nations (2023a) *AGROVOC*. Available from: https://www.fao.org/agrovoc/ [Accessed 16 September 2023].

Food and Agriculture Organization of the United Nations (2023b) *Brassica oleracea var. capitata*. Available from: https://agrovoc.fao.org/browse/agrovoc/en/page/c_9404 [Accessed 18 September 2023].

Kendal, S. & Green, M. (2019) *An Introduction to Knowledge Engineering*. London: Springer.

Lange, M.C., Lemay, D.G. & German, J.B. (2007) A multi-ontology framework to guide agriculture and food towards diet and health. *Journal of the Science of Food and Agriculture* 87(8): 1427-1434. https://doi.org/10.1002/jsfa.2832

Malik, N., Sharan, A. & Hijam, D. (2015) ‘Ontology development for agriculture domain’, *2nd International Conference on Computing for Sustainable Global Development* (INDIACom), New Delhi, 11-13 March. IEEE. 738-742.

National Geographic (2023) *The Art and Science of Agriculture*. Available from: https://education.nationalgeographic.org/resource/agriculture/ [Accessed 16 September 2023].

Ngo, Q.H., Le-Khac, N.A. & Kechadi, T. (2018) ‘Ontology based approach for precision agriculture’, *Multi-disciplinary Trends in Artificial Intelligence: 12th International Conference*, Hanoi, November 18–20, Springer International Publishing. 175-186.

Noy, N. F. & McGuinness, D. L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. Knowledge Systems Laboratory.

Pylianidis, C., Osinga, S. & Athanasiadis, I.N. (2021) Introducing digital twins to agriculture. *Computers and Electronics in Agriculture* 184: 1-25. https://doi.org/10.1016/j.compag.2020.105942

Singh, S., Shehab, E., Higgins, N., Fowler, K., Reynolds, D., Erkoyuncu, J.A. & Gadd, P. (2021) Data management for developing digital twin ontology model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 235(14): 2323-2337. https://doi.org/10.1177/0954405420978117