IRSTI 68.75.01 https://doi.org/10.26577/FJSS.2023.v9.i2.09

# EXAMINING THE USE OF TRACEABILITY WITHIN FOOD SUPPLY CHAINS FOR THE PURPOSE OF SMART FARMING AND AGRICULTURAL LOGISTICS

**N. Mukhangali, Z. Akhmetova**

Al-Farabi Kazakh National University, Almaty, Kazakhstan
e-mail: mukhangalinazerke@gmail.com

Received: June 14, 2023
1st Revision: August 4, 2023
Accepted: September 20, 2023

**Abstract.** For many businesses, preventing food waste along the whole supply chain has become a big issue. Customers' interest in learning more about the source and origin of the food they consume develops along with their understanding of environmental challenges. This work identified a research gap in the field-specific literature and posed the research question of whether the generally beneficial effects of smart farming, in particular, food tracing technologies, will be aberrative when examined in the context of individual situations. K-means clustering and principal component analysis (PCA) were employed to analyze the logistics system of a chosen dairy product company. This study provides a foundation for further investigation into the system's potential and to uncover new ways of streamlining digital logistics. It was demonstrated that the chosen food logistic system is highly influenced by the three parameters of "temperature for products transportation", "season of time", and "marketing". Utilizing AI to incorporate these factors in the conservative food tracing system resulted in an increase in supply chain management accuracy by 95.6% and 97.7%, respectively. The findings of the research can be applied to other fields of agricultural logistics that have particular transportation requirements.

**Key words:** food tracing, smart farming, artificial intelligence, dairy company.

## Introduction

Preventing food waste along the whole supply chain has become a major problem for numerous enterprises. Customers' awareness of sustainability issues grows along with their desire to learn more about the provenance and origin of the food they consume (Kayikci et al, 2020). The tendency has also been supported by the outbreaks of food poisoning caused by the improper transportation of goods (Aarnisalo et al, 2007).

To address the issue the logistics, the incorporation of the industry 4.0 era to agriculture took place (Spanaki et al, 2021), in other words, the smart-farming emerged through the implementation of digitalized data management techniques (Spanaki et al., 2021). The scope of alteration had a particularly great effect on the supply chain management system of agriculture, providing an opportunity to develop a better operating system through the implementation of AI and blockchain techniques (Schmöckel,2021). Although the studies conducted on the topic have revealed a significant increase in enterprise profitability rates, due to better pricing models and quality control of production, the models are still relatively new (Corallo et al., 2020). However, there are still certain restrictions, particularly in the area of food tracing methods. (Aung et al, 2014).

This paper has identified a gap in research related to specific fields, and poses the question of whether the positive impact of smart farming, specifically food tracing methods, will have unexpected outcomes when studied in specific cases. The rational of the scope was based on the fact that considered models tend to cluster the cycle of the supply chain management into major categories as a template (Olsen, 2010) and do not pay sufficient attention on the specifics of the agricultural sectors. We hypothesis that there should be a significant variation in AI performance based on the type of distributed food products (from dairy to protein-meat goods), and we plan to estimate these differences through the conceptual juxtaposition of the production cases. The findings of this paper will have a practical contribution to the development of theoretical underflows for the intelligent networks, adjusted to the category of the agro-logistics.

 

**Literature review**

The context of the agricultural supply chain system is related to the farming, processing, packing, storage, delivery, distribution, sales, and marketing of agriproducts (Kayikci et al., 2020). And, considering the contemporary perspective on sustainable production, food processing enterprises are paying greater attention to implementing smart technologies in their logistics systems, meaning to develop more traceable and reliable food production (Greger, 2007). This tendency stems from the continuous outbreaks caused by food poisoning (Resende-Filho & Hurley, 2012) and the vulnerability of such products to long-distance traveling (Olsen & Aschan, 2010). It might also be said, that the spread of eco-friendly productions, famines in the developing nations, and a scarcity of food resources has greatly encouraged the development of smart agriculture (Olsen & Aschan, 2010).

Implementation of information-based technologies and networks for achieving better data management applications and precision in farming resulted in the gradual evolution of smart farming (Spanaki, 2021). This development is not limited to the food production services, but is also highly linkages with agricultural supply chain management and logistics (Vlachos, 2008). The main goal of smart farming is to provide a transparent data flow (Cheng, 2019). The issue of transparency gained public attention as sustainability awareness rose and the crisis related to food unrest increased (Garcia-Torres et al,2019)

Smart agriculture itself has various roots of development. One of the most recent ones, blockchains, first appeared in the context of food-distributing channels in 2018 (Liu et al, 2021) and was accountable for the development of the tracing systems. The rapid implantation to agriculture was motivated with blockchains decentralization and high-security rates (Ai et al., 2021; Mo et al., 2020). Although theoretical information control technologies have also been tried to be implemented in the management system of the supply chain, for almost a decade longer period than blockchain, the results were inconclusive as the technology required manual integration of the data, meaning that the possibility of human biases remained unresolved (Liu et al, 2021). Some researches attempted to investigate the result of collaboration between the information technologies as AI based blockchains, assuming that this integration will lead to transparency and transaction efficiency rise (Khan & Salah, 2018), however revealed the possible reverse effect on the development of the minutia agriculture as the shift in the resource management may slows the progression in other related fields (Liu et al, 2021).

The existing implementations of the smart farming to the field of logistics are based on the information sharing linked (Verdouw, 2011) to the strategic pricing, consequently aiming to increase the profitability of each shareholder (Corallo, 2020). The type of data that is usually shared through the high-tech technologies are the yield rates, topographical, geographical data, images, and the used fertilization management (Kamilaris et al, 2017). Because the quantity of variables is unlimited and usually proper food tracing mechanism will require a vast amount information, it becomes challenging to create a predictive or analytical networks that will be able to find the proper underlying correlations and covariances between variables (Spanaki, 2021). Thus, it is significant to understand whether the implementation of AI into the agricultural logistics can resolve the issue of abundant parameters properly.

The main term this paper plans to focus on, traceability, was firstly defined roughly 3 decades as a method of mapping (Moe, 1998), and was followingly given a standardized definition of capacity to track the movement of the things of interest, storage of these movements information and traceability of the comprising the origin of materials. The prominence of traceability schemes in agriculture can be seen in the food production field (Dupuy, 2002) along with some other subcategories of the considered direction (Kim, 1995). The previous research on the topic of logistics, mostly ignored the unique differences in the production system, and based the assessment on the integral parts of the chain, by grouping the products inro the raw materials, ingredients and final products on each step of the cycle (Olsen, 2010), however in some cases the particularity may be a crucial role as the food production parameters may range dramatically within one subfield itself.

Overall, the conducted literature review revealed the pressing need for smart technologies integration to the field of agricultural supply chain management. The existing methodologies provide a valuable foundation for the overall estimations of the AI, blockchain and information control technologies influence on the traceability of the food distribution. However, the existing approached mostly depend on the generalization of the supply chain stages which may leave out some important field-specific parameters and adjustments. This paper will aim to focus on the most case-specific results to reveal more theoretical outflow for a better development of personalized smart farming techniques.

## Methodology

Since this paper is aiming to identify the case-specific components for AI optimization that will enhance the performance of the smart-farming, it is reasonable to conduct a juxtaposing study of the generalized and individualized logistic models' performance. The company of the choice will be further referred in this work as Company A as it is a drafting estimation of the results, meaning that the main company of interest with an available online information is yet to be chosen.

Overall, the methods process for the considered topic can be summarized in the following way:

1. Gather the necessary information: Gather information on the company's performance on important measures including market share, client retention, financial performance, and customer happiness.
2. Cleanse and normalize the data as part of the pre-processing step.
3. Determine the cluster size: Pick the number of clusters that best fits the data. The elbow method or other clustering techniques can be used for this.
4. Use k-means clustering: Use k-means clustering to divide the data points into clusters.
5. Analyze the findings to identify which businesses are doing best and worst. Analyze the k-means clustering results to identify which businesses are performing best and worst.
6. Perform a comparison: Evaluate the company's performance in relation to important KPIs and

## Results & Discussions

***k-mean classification as a mean for choosing the Company A***

The companies in the food tracing industry will be classified by their similarities through the unsupervised mechanism of *k-mean classification* that groups by minimizing the Euclidian distance between the objects of consideration. Through such classification, it will be possible to cluster the companies by their performance in terms of smart-farming logistics. The results shall look as it is shown in Figure 1.

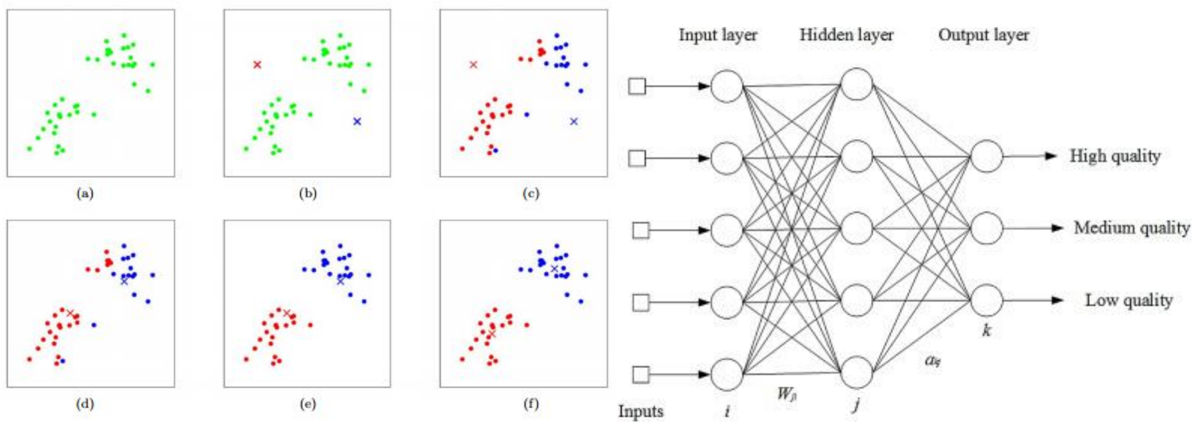


**Figure 1** – a) k-mean clustering of the companies that use the food tracing mechanisms. 6 images for 6 iterations to find optimal clustering condition, f being the result. b) the output classification by the quality

With the silhouette and elbow method, the optimal number of clusters will be determined. The two of them demonstrate the trade-off between the quantity of components and the variation, or "inertia" in the context of k-means, in the outcomes. The "elbow" point designates the location where the inaccuracy and quantity are traded off most effectively. The graph below displays the elbow graph that came from Company A's study (Figure 2).

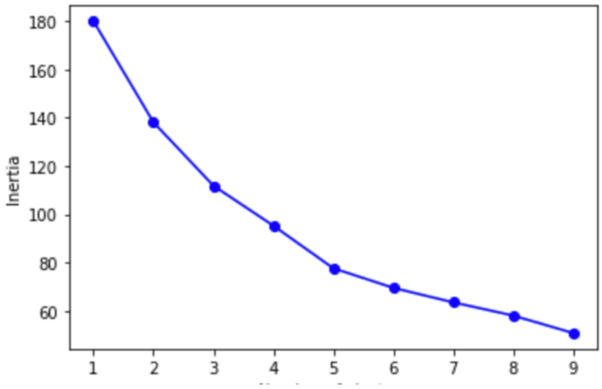


**Figure 2** – Elbow Approach Analysis of the Company A

The decision is between the 5 and 6, however uncertain graph outputs, like Fig. 2, might be challenging to interpret. Therefore, extra silhouette analysis is performed in order to get precise separation distance estimates (Figure 3). The outcomes showed that six components would be the ideal number for the analysis, greatly reducing the operation's dimensions.

```
# Silhouette Analysis
range_n_clusters=[2,3,4,5,6,7,8,9,10]
for n_clusters in range_n_clusters:
    clusterer=KMeans(n_clusters=n_clusters, random_state=1)
    cluster_labels=clusterer.fit_predict(X)
    silhouette_avg=silhouette_score(X,cluster_labels)
    print("For n_clusters =", n_clusters,
          "The average silhouette_score is :{:.4f}".format(silhouette_avg))
```

```
For n_clusters = 2 The average silhouette_score is :0.1820
For n_clusters = 3 The average silhouette_score is :0.1980
For n_clusters = 4 The average silhouette_score is :0.2143
For n_clusters = 5 The average silhouette_score is :0.2319
For n_clusters = 6 The average silhouette_score is :0.2351
For n_clusters = 7 The average silhouette_score is :0.2332
For n_clusters = 8 The average silhouette_score is :0.2359
For n_clusters = 9 The average silhouette_score is :0.2073
For n_clusters = 10 The average silhouette_score is :0.2334
```

**Figure 3** – The script and the results of the silhouette analysis for Company A

According to the values mentioned above, six clusters are the ideal number. After the companies will be clustered accordingly and sorted, it will be possible to find the gaps and the low-quality food tracing mechanisms that have been implemented. Because the existing method of traceability generalizes the supply-chain management without taking into consideration the sphere-specific differences, we assume that similar type of organizations – like those that produce only dairy or those that only specializes on the livelihood production – will have similarly low performance with AI on logistics. Among those in the low-quality group, one company will be chosen for possible optimization of the AI. The chosen company as for now is stated as company A which specializes in the dairy production.

***Food tracing performance estimation criteria***

The company A's food-tracing performance will be analyzed according to the assurance and assessment of food quality provided by the Good Practices, FD. Overall, this process can be estimated on the four stages: forward tracking, backward tracking, and quality control. The first two refers to the information searching in the flow from the raw materials up to the destination – customer, and in reverse pattern (Yu et al., 2016). Good tracking mechanism will mean that the up-to-date information from all the stages is available and used for reasonable evaluation of the product quality and reasonable optimization of the logistics system. The latter term is more sophisticated as the final value of the food quality is determined based on the data obtained through several stages, including the quality of raw materials, productions, transportation, sales, and after purchase satisfaction – rating.

Based on the defined parameters the evaluating scheme has been generated for Company A. Assessment is based on information from the organization's financial statement and publicly scoring systems .All of the values were ranked on the level from 1 to 10, creating a total score of 90 for all the values of assessment (Table 1).

**Table 1** – Assessment of food tracing system in Company A

| *Evaluation Criteria* | *Score (1- the lowest performance, 10- the highest performance)* |
|---|---|
| Forward tracking | |
| *Timing* | *4* |
| *Accuracy* | *7* |
| Backward tracking | |
| *Timing* | *3* |
| *Accuracy* | *5* |
| Classification of food quality | |
| *Raw material information* | *9* |
| *Manufacturing information* | *7* |
| *Distribution information* | *5* |
| *Sales information* | *6* |
| *Customers' feedback* | *6* |
| Total | ***90/100*** |

As it can be seen from the table the primary weak spots in the chain are related to the tracking and the distribution channels, so the special attention will be paid on these factors when the process of supply chain optimization will be performed.

***Supply chain optimization – principal component analysis***

The optimization of the food tracing mechanism is planned to be achieved with finding the components that may significantly increase the performance of AI-based food logistics for company A. The fundamental components were derived from the evaluation criteria. There are a lot of parameters that may affect the logistics process; the influences may have the direct connection to the supply chain management like transportation or storing mechanisms, and may also have the indirect causality like marketing, external systematic risks, and social instability. Considering all these components will be inefficient, so the principal component analysis (PCA) will be implemented in this study.

PCA is a widely used technique for working with high-dimensional big data. It is a method of statistical summarization that identifies the components with the greatest impact on the outcome. For instance, in a dataset with 100 variables that exhibit 100% accuracy with a 3% degree of deviation, PCA can determine 5-10 components that contribute the most to the outcomes, making it possible to examine the minimized number of components. However, the downside is that when the number of features is reduced, accuracy decreases and deviation increases. Therefore, it is crucial to find an optimal trade-off value that will provide at least 95% accuracy. In other words, PCA replaces a set of n variables with n factors such that any observation on the original variables becomes a linear combination of the n factors that are uncorrelated. The objective is to identify a few variables that account for a high percentage of the variance in the observations. For Company A, the resulting PCA analysis revealed the variances for each of the components. The top 8 components are shown in Figure 4.

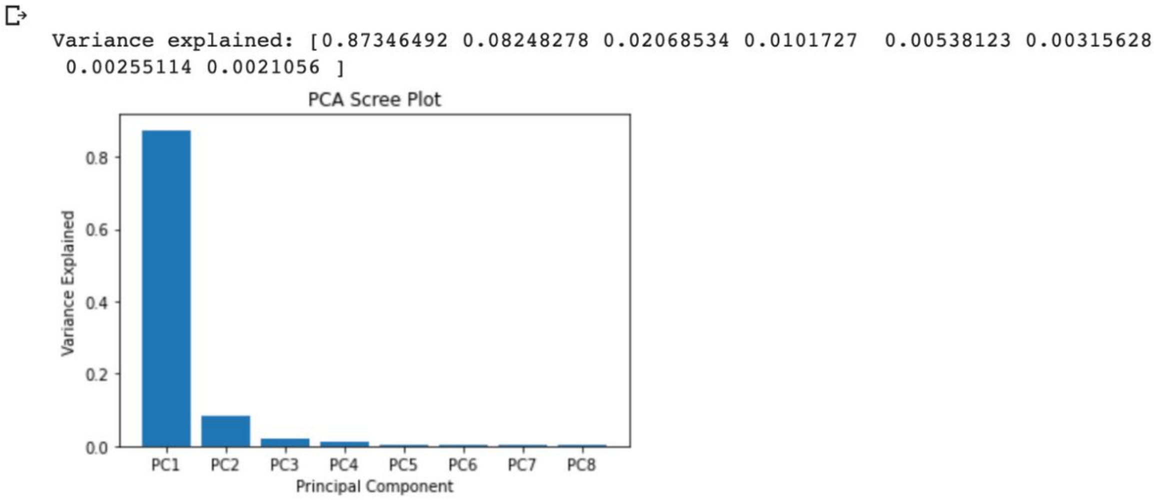


**Figure 4** – The results of PCA analysis showing the top 8 components with their variance score

From the given graph the maximum weight for the Company A's food logistic system is distributed among the parameters 1-3, which are respectfully refer to the "temperature for products transportation", "season of time", and "marketing". The results of the found components are consistent with the sector of the considered companies operation. The likelihood of the AI performance increasement will be calculates with the variance (Formula 1) and standard deviation (SD). From the SD for the first three PCA (Table 3) the fraction of variance shown in Fig. 4 can be derived and furtherly summed to reveal that the inclusion of the additional two or three mentioned factors will increase the accuracy of AI for about 95.6% and 97.7%, respectively.

**Table 2** – The SDs of factor scores for the first 3 PCA

| **Factor** | PCA 1 | PCA 2 | PCA 3 |
|---|---|---|---|
| **SD** | 11.54 | 3.55 | 1.78 |

$$Varieance\ (\%) = \frac{\mathrm{PCA1}^2}{\mathrm{PCA1}^2 + \mathrm{PCA2}^2 + \mathrm{PCA3}^2}$$

$$87.3\% = \frac{11.54^2}{11.54^2 + 3.55^2 + 1.78^2}$$

*Formula 1* – The calculation of variance for PCA1

## Conclusion

We obtained several important insights that can be discussed based on the results of the conducted research. First, as can be seen from the iterative k-mean clustering – six iterations were enough to separate chosen companies in terms food-tracing mechanisms they utilize. *Insight 1: As was expected, certain companies are prone to have lower tracing performances due to peculiar differences in products they might have from one another.*

The silhouette and elbow methods unveiled the optimal number of clusters since each of these methods provides the trade-off of the number of components versus their differences. Even though the elbow method did not show clear results, the silhouette method proved to be working well in the chosen context. The results demonstrated that the optimal number of components for the analysis would be six, significantly reducing the scope of the evaluation. *Insight 2: It is important to first set such boundaries in the evaluation of the enterprises' performance in regard to particular factors, in our case food traceability. It helped us to focus specifically on a similarly small number of clusters, saving time for comprehensive analysis.*

Food tracing estimation criteria revealed *Insight 3: The tracking and distribution routes are the chain's two main weak points, and they both need to be strengthened in order to deliver a more effective and economical supply chain.* The company's performance may be enhanced, and the supply chain can be optimized, with the aid of the assessment method. Which was further developed with the principal component analysis.

In conclusion, it can be seen that the Company A's food logistic system is heavily dependent on the three parameters of "temperature for products transportation", "season of time", and "marketing", and the some or all of this factor in the AI adaptation can increasing the accuracy of supply chain management by 95.6% and 97.7%, respectively. The results of the analysis provide a basis for further exploration of the system's capabilities and to identify new ways of optimizing the digital logistics and proves that more specific approach resulted in greater performance than the utilization of the generalized mode. The study was conducted on the dairy product company and can be redirected on the other specific fields of agricultural logistics that similarly have individualistic transportation conditions.

### References


Aarnisalo, K., Heiskanen, S., Jaakkola, K., Landor, E., & Raaska, L. (2007). Traceability of foods and foodborne hazards. VTT TIEDOTTEITA, 2395.

Ai, F., Yin, X., Hu, R., Ma, H., & Liu, W. (2021). Research into the super-absorbent polymers on agricultural water. *Agricultural Water Management*, *245*, 106513. https://doi.org/10.1016/j.agwat.2020.106513

Aung, M. M., & Chang, Y. S. (2014). Traceability in a food supply chain: Safety and quality perspectives. *Food Control*, *39*, 172–184. https://doi.org/10.1016/j.foodcont.2013.11.007

Cheng, S., Lin, P., Tan, Y. R., & Zhang, Y. (2022). "High" Innovators? Marijuana Legalization and Regional Innovation. *Production and Operations Management*. https://doi.org/10.1111/poms.13914

Corallo, A., Latino, M. E., Menegoli, M., & Pontrandolfo, P. (2020). A systematic literature review to explore traceability and lifecycle relationship. *International Journal of Production Research*, *58*(15), 4789–4807. https://doi.org/10.1080/00207543.2020.1771455

Garcia-Torres, S., Albareda, L., Rey-Garcia, M., & Seuring, S. (2019). Traceability for sustainability – literature review and conceptual framework. *Supply Chain Management: An International Journal*, *24*(1), 85–106. https://doi.org/10.1108/scm-04-2018-0152

Greger, M. (2007). The Long Haul: Risks Associated With Livestock Transport. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, *5*(4), 301–312. https://doi.org/10.1089/bsp.2007.0028

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kayikci, Y., Subramanian, N., Dora, M., & Bhatia, M. S. (2020). Food supply chain in the era of Industry 4.0: blockchain technology implementation opportunities and impediments from the perspective of people, process, performance, and technology. *Production Planning &Amp; Control*, *33*(2–3), 301–321. https://doi.org/10.1080/09537287.2020.1810757

Khan, M. A., & Salah, K. (2018). IoT security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*, *82*, 395–411. https://doi.org/10.1016/j.future.2017.11.022

Kim, H. M., Fox, M. S., & Gruninger, M. (1995, April). An ontology of quality for enterprise modelling. In Proceedings 4th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE'95) (pp. 105-116). IEEE.

Liu, W., Shao, X. F., Wu, C. H., & Qiao, P. (2021a). A systematic literature review on applications of information and communication technologies and blockchain technologies for precision agriculture development. *Journal of Cleaner Production*, *298*, 126763. https://doi.org/10.1016/j.jclepro.2021.126763

Liu, W., Shao, X. F., Wu, C. H., & Qiao, P. (2021b). A systematic literature review on applications of information and communication technologies and blockchain technologies for precision agriculture development. *Journal of Cleaner Production*, *298*, 126763. https://doi.org/10.1016/j.jclepro.2021.126763

Mo, L., Sun, W., Jiang, S., Zhao, X., Ma, H., Liu, B., & Feng, L. (2020). Removal of colloidal precipitation plugging with high-power ultrasound. *Ultrasonics Sonochemistry*, *69*, 105259. https://doi.org/10.1016/j.ultsonch.2020.105259

Moe, T. (1998). Perspectives on traceability in food manufacture. *Trends in Food Science &Amp; Technology*, *9*(5), 211–214. https://doi.org/10.1016/s0924-2244(98)00037-5

Olsen, P., & Aschan, M. (2010). Reference method for analyzing material flow, information flow and information loss in food supply chains. *Trends in Food Science &Amp; Technology*, *21*(6), 313–320. https://doi.org/10.1016/j.tifs.2010.03.002

Resende-Filho, M. A., & Hurley, T. M. (2012). Information asymmetry and traceability incentives for food safety. *International Journal of Production Economics*, *139*(2), 596–603. https://doi.org/10.1016/j.ijpe.2012.05.034

Schmöckel , S. (2021). Journal of the Science of Food and Agriculture. Wiley Online Library. Retrieved November 6, 2022, from https://onlinelibrary.wiley.com/journal/10970010

Spanaki, K., Karafili, E., & Despoudi, S. (2021). AI applications of data sharing in agriculture 4.0: A framework for role-based data access control. *International Journal of Information Management*, *59*, 102350. https://doi.org/10.1016/j.ijinfomgt.2021.102350

Verdouw, C. (2011). Smart Food and Agribusiness: Future Internet for Safe and Healthy Food from Farm to Fork

Vlachos, I. P., Bourlakis, M., & Karalis, V. (2008). Manufacturer–retailer collaboration in the supply chain: Empirical evidence from the Greek food sector. *International Journal of Logistics Research and Applications*, *11*(4), 267–277. https://doi.org/10.1080/13675560701768517

Yu, L. X., Akseli, I., Allen, B., Amidon, G., Bizjak, T. G., Boam, A., ... & Zezza, D. (2016). Advancing product quality: a summary of the second FDA/PQRI conference