IDENTIFYING THE OPPORTUNITIES FOR THE DESIGN OF DIGITAL PLATFORMS: A TOPIC MODELLING APPROACH

VEENA KANNAN[1], NORA FTEIMI[2], SAJI K MATHEW[1] & FRANZ LEHNER[2]

[1] Indian Institute of Technology Madras, Department of Management Studies, Chennai, India, e-mail: veenakannan1994@gmail.com, saji@iitm.ac.in
[2] University of Passau, School of Business, Economics and Information Systems, Passau, Germany, e-mail: nora.fteimi@uni-passau.de, franz.lehner@uni-passau.de

Abstract Aquaculture is one of the fast-growing food-producing agriculture subsectors. However, the digital infrastructures developed in aquaculture are self-organising platforms i.e. they do not rely on a centralized intermediary for monitoring, coordinating activities or for overseeing transactions. Hence, the main objective of this research paper is to identify the challenges farmers face in an entire supply chain for designing a digital platform for the aquaculture domain. The main problems faced by the farmers include water quality issues, disease outbreak, lack of proper information regarding suitable insurance policies etc. We have identified eight such issues that the farmers face in an entire harvest period and also prioritized them. The results from our study could be used for the further advancement of an integrative perspective in the design and implementation of the digital platform for aquaculture.

1   Introduction

Digital innovation includes carrying out new combinations of digital and physical components in order to produce new digital infrastructures (Yoo, Ola, & Lyytinen, 2010). Digital platforms are built and integrated on top of digital infrastructures. We define digital platforms as "*a set of digital resources— including services and content—that enable value-creating interactions between external producers and consumers*" (Parker, Van Alstyne, & Choudary, 2016). Digital platforms transform the transaction logic as they ease transactions between distinct supply chain stakeholders rather than handling the entire supply and logistics chain on their own (Hänninen, Smedlund, & Mitronen, 2018). For example, in the case of the automotive industry, subsystems (e.g., voice assistants or navigation systems) are becoming digitized independently but connected via vehicle-based software architectures. Hence, firms from other industries can also develop and integrate new products and services with the computing platform of the automotive industry (Henfridsson & Lindgren, 2010). However, in many industries, especially in agriculture (Bookie & Duncombe, 2019) and its allied subsectors like aquaculture (Mathisen, Haro, Hanssen, Bjork, & Walderhaug, 2016) there is a lack of centralized approach for distribution of content (e.g., tailored information regarding disease prediction, water quality parameters etc.). The digital infrastructures in the aquaculture domain are self-organizing platforms i.e. they do not rely on a centralized intermediary for monitoring, coordinating activities or for overseeing transactions between farmers and exporters (Forte, Larco, & Bruckman, 2009). Recent research shows that 80% of such platforms face challenges in ensuring content integrity which can undermine its survival (Tiwana & Bush, 2014). But, designing a digital platform upfront for a particular industry is challenging, as platforms change the power structure and relationship between different stakeholders (De Reuver, Sørensen, & Basole, 2018). Therefore, the main objective of this paper is to investigate how data-driven approaches can affect the design of collaborative digital platforms research from self-organizing platforms in the special case of aquaculture. In order to improve the design, we have to identify the problems faced by the stakeholders in each phase of the aquaculture supply chain (e.g. disease prediction, water quality monitoring etc.). Therefore, we would like to contribute to research by identifying and analyzing such occurring problems from prior literature.

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

219

According to the reports of Food and Agricultural Organization (FAO) in 2014, aquaculture will contribute majorly in the future to food security and adequate nutrition for the growing world population which is expected to reach 9.7 billion by the year 2050 (Subasinghe, Curry, Mcgladdery, & Bartley, 2003). Moreover, FAO also reports the challenges aquaculture is facing over the decades such as combating diseases, brood stock improvement and domestication, development of efficient feeding mechanisms, water-quality management etc. A decision support digital platform supporting the entire aquaculture supply chain has not received due attention in literature even though there is a considerable amount of literature for fisheries (Mathisen et al., 2016). Therefore, we integrate a list of identified factors from prior literature for designing a digital platform for the aquaculture industry.

This paper proceeds as follows: related work is outlined in section 2. Section 3 describes the research objective. In sections 4 and 5, we discuss the methodology and main results respectively and ends up in section 6 with a conclusion.

2   Related Work

A cursory examination of the literature on digital platforms reveals the diverse orientation of studies in this area. For instance, some studies were focusing on platform for healthcare, and some for platforms for energy informatics. In order to identify the key empirical studies related to the design of digital platforms, we conducted a systematic literature review (Webster & Watson, 2002) following a process of searching, filtering and classifying related papers. As the research on digital platforms started to appear in IS journals in the year 2002 (Asadullah, Faik, & Kankanhalli, 2018), we searched for articles between the period 2002 and 2019. We conducted a search in the databases of AISeL, IEEE, EBSCO and Google Scholar for accessing relevant journal publications and conference proceedings, using keywords "multisided platforms", "digital platforms", "two-sided markets" and their combinations and obtained 1000 hits. The papers were shortlisted first based on the title, after which we further shortlisted based on abstract relevance. Thereby, we included only papers dealing with the design factors of digital platforms in different industries and also research commentaries focusing on the design and governance of digital platforms (e.g. De Reuver et al., 2018). Research papers dealing with other topics like pricing, competition in digital platforms were excluded. Based on these criteria, a summary of 40 relevant articles resulted, which we used as base for our

analysis. The requirements of digital platforms along with the key stakeholders collaborating in the platform and some studies in the respective domains are mentioned in Table 1.

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

221

Table 1: Digital platform requirements in various domains

Digital platform domain	Platform requirements	Stakeholders	References
Agriculture	Reduction of information asymmetry, higher bargaining power and access to profitable markets.	Farmers, Buyers	Banker et al., (2011), Goyal, (2010), Jha, Pinsonneault, & Dubé, (2016)
Electricity	Competitive gaming platform have to enable dynamic electricity trading.	Electricity providers and subscribers.	Ketter, Peters, Collins, & Gupta, (2016)
Finance	Crowdfunding platform has to enable entrepreneurs to raise funds quickly	Entrepreneur, Funding agencies	Abhishek, Geng, Li, & Zhou, (2017), Hernando, (2007)
Healthcare	Patients data have to be managed in electronic health records, better healthcare assistance, higher productivity has to be ensured	Doctors, Nurses, Hospital administration and patients	Rodon, (2018), Agarwal et al., (2010), Jones, (2014)

3   Research Objective

The majority of studies performed in the Information Systems (IS) domain and its reference disciplines focuses on pricing (Rochet & Tirole, 2003) while neglecting technical, social and strategic aspects of the platforms (Pettigrew, Woodman, & Cameron, 2001). Moreover, the recent literature review by De Reuver et al. (2018) calls for more research on data-driven approaches to inform the design of digital platforms. Even though the decision support platform has been well studied in fisheries, there are lack of studies for the equivalent research in aquaculture (Mathisen et al., 2016). In order to fill these knowledge gaps, the main research objective of this study as follows:

How can data-driven approaches affect the design of collaborative digital platforms research from self-organizing platforms in the special case of the aquaculture domain?

4   Research Methodology

More than 80 percent of data today is stored in unstructured formats like audio, video, text etc. making it a difficult task to search, organize, synthesize and understand this huge corpus of information (Debortoli, Müller, & Junglas, 2016). To capture the problems faced by farmers in the aquaculture supply chain, we have analyzed the abstracts of leading aquaculture engineering journals and conference proceedings. There are supervised and unsupervised text mining approaches to classify textual data. Since the design choices of a digital platform are unclear, we rely on an unsupervised learning approach (e.g. Vidaurre, Kawanabe, Bünau, Blankertz, & Müller, (2011)) to allow the algorithm the autonomous identification of challenges faced by farmers in the text. The unsupervised approach allows us to discover the latent topics from the written abstracts. Latent Dirichlet Allocation (LDA), is one of the most widespread techniques used in the IS domain in order to identify common topics and their distributions (e.g. Rodriguez & Piccoli, 2018) from textual data (Eickhoff & Neuss, 2017). In the following, we provide more details on our dataset and the text mining approach used.

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

223

4.1   Data Sample

Research on digital infrastructures for aquaculture is relatively recent and therefore in an early progress stage. To get a broader picture of problems addressed by these infrastructures at different aquaculture supply chain phases, we collected articles from different databases such as EBSCO, Google Scholar and Scopus. We searched using the keywords "digital aquaculture", "machine learning and aquaculture", "artificial intelligence and aquaculture" and their combinations. To ensure quality, we focused on journals such as Sensors, Computers and Electronics in Agriculture, Aquaculture Engineering, Aquaculture, Reviews in Aquaculture and leading conference proceedings of IEEE, ACM. As a result, 50 documents over a period of 20 years were finally included consisting of 35 journal papers and 15 conference proceedings. After selecting the documents, we transformed the PDF files in image format to text format. This transformation was necessary for the subsequent text mining analysis. We focused the analysis on the abstracts and relevant parts of the introduction section, as inspecting the documents revealed that these parts already provided detailed information about the problems in the aquaculture supply chain and therefore fits well the purpose of this study.

4.2   Topic Modelling Approach for Data Analysis

Topic modelling algorithms are statistical methods for understanding latent topics inherent in text documents to help researchers summarize and interpret collected information along with topic labels (Blei, 2012). In our approach, feature extraction technique of Latent Dirichlet Allocation (LDA) and Rapid Automatic Keyword Extraction (RAKE) was applied using the statistical software programming software "R". LDA is one of the generative probabilistic algorithm commonly used in text mining and topic labelling (Blei, 2003). As an unsupervised learning algorithm, each document is treated as a bag of words to discover the latent topics from the distribution over words. During pre-processing, the data corpus was first tokenized by splitting sentences into words, removing punctuation, white space and numbers. After that, we removed stop words and performed stemming and lemmatization based on the WordNet database (Fellbaum, 1998). Subsequently, LDA was performed to determine the topics and their associated words. Prior to the analysis, we infer the optimal number of topics during an iterative procedure, which resulted in a number of eight topics that best represented the problems faced by farmers

during aquaculture cultivation. Topic labels can be compiled by grouping topic words to common-higher level themes. Human judgement can be used for performing topic labelling (Syed & Dhillon, 2015; Shi, Lee, & Whinston, 2016). Therefore, labelling of the problems faced by farmers and grouping into design features for platforms were performed by two independent researchers with extensive experience in digital platforms.

5   Results and Discussion

We computed the top words for each topic (figure 1 to figure 4) and also sorted topics according to their importance within the entire document collection (figure 5) using the gama function and theta function in R respectively. We also checked the document to topic probability to ensure the validity of the results.
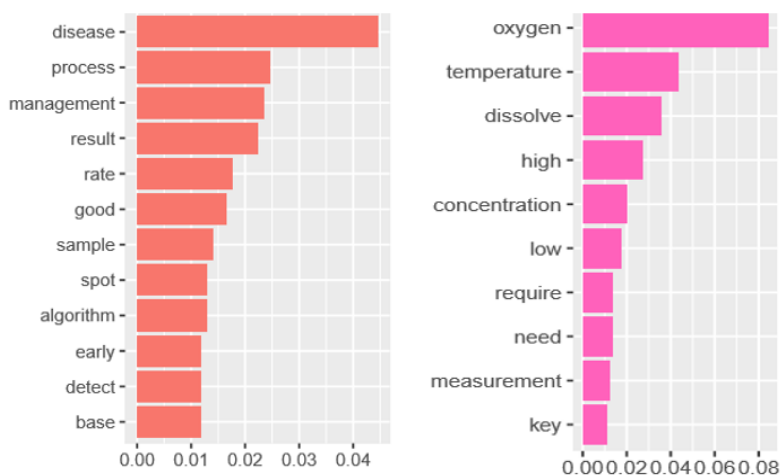

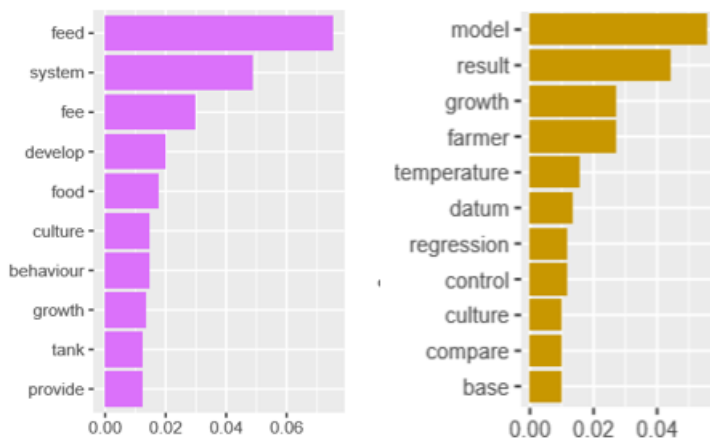**Figure 1: #T1: Disease prediction, #T2: Dissolved oxygen concentration prediction**

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

225

Figure 2: #T3: Feed parameters monitoring, #T4: Growth prediction
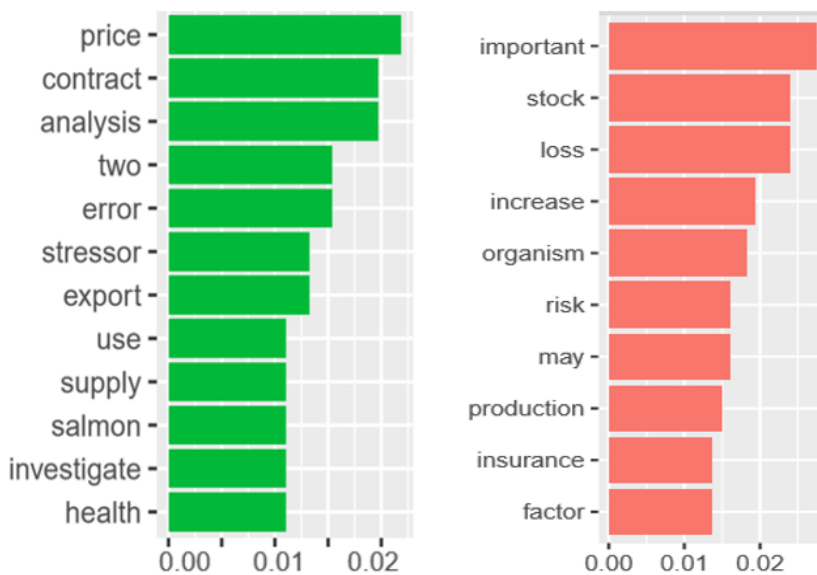
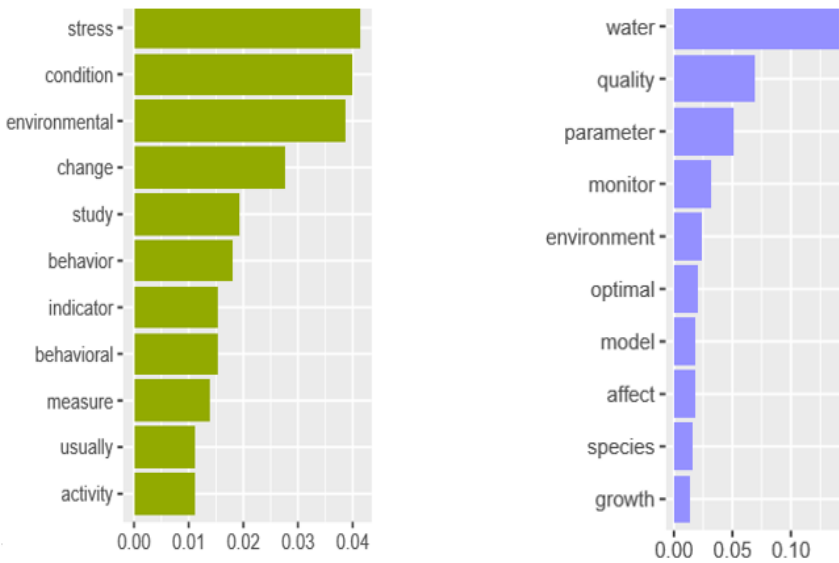**Figure 3: #T5: Market price prediction, #T6: Insurance policies available**

Figure 4: #T7: Behaviour monitoring & alerting, #T8: Water quality monitoring & alerting

As figure 5 shows one of the main problems faced by the farmers is the monitoring of water quality parameters (#T8). Water quality is an important aspect of aquaculture harvest and if any of the parameters are not at the optimum level may affect the animal health badly and can cause loss to the farmers (Piplani et al., 2015). In many of the aquaculture harvesting practiced in the developing countries, the farmers monitor these parameters manually and perform lab tests weekly. These traditional methods are amongst others time consuming, difficult for decision-making. Along with water quality monitoring and alerting the core of the platform can employ machine learning and artificial intelligence techniques for prediction of growth (#T4), disease (#T1), dissolved oxygen concentration (#T2) and monitoring of the behavior of the animal (#T7)(Yu, Leung, & Bienfang, 2006). The right amount of feed can also be automated, by the core of the platform as this accounts for major cost in the aquaculture supply chain (#T3). Along with the core, the main complements that need to support the digital platform from our analysis are insurance agents (#T6) and suitable exporters by predicting the market price (#T5). As aquaculture involves high risk as it requires estimating different parameters, the platform should be equipped to provide reliable farm level data to show proof for the insurance agencies in case of harvest loss (Secretan, 2008). Moreover, the farmers in traditional methods rely on intermediaries to find out suitable exporters. However,

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

227

with the digitalization of the aquaculture supply chain, the farmers could provide evidence for the quality of the product and bargain for higher value with a wide variety of exporters and choose the best price offered.
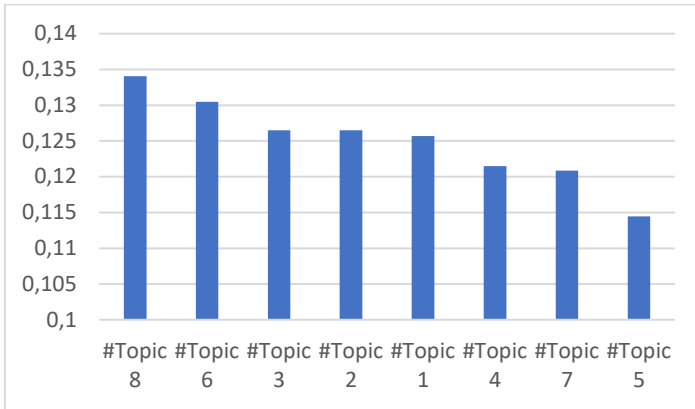

Figure 5: Topic Ranking

6   Conclusion and Limitation

In this study, we examined 50 digital aquaculture research works by applying machine learning approach of topic modelling. Our aim was to examine the factors that need to be considered while designing a collaborative digital platform from multiple self-organizing platforms in the case of the aquaculture supply chain. While there has been research on digital infrastructure in aquaculture in recent years, there is a paucity of studies that focus on the analysis, integration and temporal comparison of these infrastructures addressing problems faced by farmers in different phases of the aquaculture supply chain. The main stakeholders we identified from our data analysis for the digital platform are farmers, exporters and insurance agents.

The focus of our research was to understand the challenges the farmers face in an entire aquaculture supply chain. Our results bring out eight challenges in the aquaculture domain that have to be considered while designing a digital platform. Water quality monitoring, insurance policies available and feed parameters monitoring are some of the main problems that have to be addressed. Water quality parameters have to be controlled appropriately in the optimum range to increase the

fish growth rate and to reduce the outbreak of diseases (Stigebrandt, Aure, Ervik, & Hanson, 2004). After water quality, the second most important challenge identified in our research is suitable insurance policies. As the stock in aquaculture is grown in water, it is prone to unique risks and hazards unlike other industries (Secretan, 2008). Therefore, these are the two most important factors that have to be considered while designing a digital platform for aquaculture. Our results pinpointed that quantitative analytical methods such as LDA can be used for the analysis of qualitative data as that of research papers to get a bigger picture and insights for understanding the problems they address in general.

However, this study also has some limitations. The first limitation is the limited size of the dataset for topic modelling. However, as we included larger parts of the papers, our focus was not on quantity but on obtaining a detailed topic analysis of our specific text segments to identify concrete design factors for digital platforms. Future research can derive more insights into the problems faced by the farmers by including further textual data sources such as newspaper articles, data available on social media, blogs etc. Secondly, negation detection has not been considered in the algorithm. However, as we carefully inspected some text parts in advance during data preparation, we are confident that this limitation will not affect the final results substantially. Thirdly, with regard to methodology, topic modeling was the only method applied in this study. Future research can compare our results with the results from other methodologies like expert surveys for further verification. Moreover, future works could also focus on using design science principles to build a prototype version of the digital platform by considering the outputs from this research. Topics pertaining to the problems faced by the farmers in special contexts such as that of developing countries will be also interesting. The prototype can be verified to obtain validation of the literature results through direct interaction with farmers and expert opinion acquisition.

References

Abhishek, V., Geng, D., Li, B., & Zhou, M. (2017). When the Bank Comes to You: Branch Network and Customer Omni-Channel Banking Behavior. Ssrn, 1–32. https://doi.org/10.2139/ssrn.3033988

Agarwal, R., Gao, G. G., DesRoches, C., & Jha, A. K. (2010). The digital transformation of healthcare: Current status and the road ahead. Information Systems Research, 21(4), 796–809. https://doi.org/10.1287/isre.1100.0327

Veena Kannan, Nora Fteimi, Saji K Mathew and Franz Lehner:
Identifying the Opportunities for the Design of Digital Platforms: A Topic Modelling Approach

229

Asadullah, A., Faik, I., & Kankanhalli, A. (2018). Digital Platforms: A Review and Future Directions. PACIS 2018 Proceedings. https://doi.org/10.1017/CCOL9780521862387.020

Banker, R., Mitra, S., & Sambamurthy, V. (2011). The Effects of Digital Trading Platforms on Commodity Prices in Agricultural Supply Chains. MIS Quarterly, 35(3), 599–611. https://doi.org/10.2307/23042798

Blei, D. M. (2003). Latent dirichlet allocation. Journal of Machine Learning Reseaarch, 3, 993–1022.

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77–84. https://doi.org/https://doi.org/10.1145/2133806.2133826

Bookie, E., & Duncombe, R. (2019). The role of digital platforms in disrupting agricultural value chains in developing countries. International Conference on Social Implications of Computers in Developing Countries, 231–247. https://doi.org/0.1007/978-3-030-18400-1_19

De Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. Journal of Information Technology, 33(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

Debortoli, S., Müller, O., & Junglas, I. (2016). Text Mining For Information Systems Researchers : An Annotated Topic Modeling Tutorial. Communications of the Association for Information Systems, 39(7), 110–135. https://doi.org/10.17705/1CAIS.03907

Eickhoff, M., & Neuss, N. (2017). Topic Modelling Methodology : Its Use In Information Systems and other Managerial Disciplines. Proceedings of European Conference on Information Systems, 2017, 1327–1347.

Fellbaum, C. (1998). A semantic network of English verbs." WordNet: An electronic lexical database. An Electronic Lexical Database, 3, 153–178.

Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia Governance. Journal of Management Information Systems, 26(1), 49–72. https://doi.org/https://doi.org/10.2753/MIS0742-1222260103

Goyal, A. (2010). Information , Direct Access to Farmers , and Rural Market Performance in Central India. American Economic Journal: Applied Economics, 2(3), 22–45. https://doi.org/10.1257/app.2.3.22

Hänninen, M., Smedlund, A., & Mitronen, L. (2018). Digitalization in retailing: multi-sided platforms as drivers of industry transformation. Baltic Journal of Management, 13(2), 152–168. https://doi.org/10.1108/BJM-04-2017-0109

Henfridsson, O., & Lindgren, R. (2010). User involvement in developing mobile and temporarily interconnected systems. Information Systems Journal, 20(2), 119–135. https://doi.org/10.1111/j.1365-2575.2009.00337.x

Hernando, I. (2007). Is the Internet delivery channel changing banks ' performance ? The case of Spanish banks. Journal of Banking and Finance, 31, 1083–1099. https://doi.org/10.1016/j.jbankfin.2006.10.011

Jha, S. K., Pinsonneault, A., & Dubé, L. (2016). The Evolution of an ICT Platform-Enabled Ecosystem for Poverty Alleviation: The Case of eKutir. MIS Quarterly, 40(2), 431–445. https://doi.org/10.25300/MISQ/2016/40.2.08

Jones, M. (2014). A Matter Of Life And Death: Exploring Conceptualizations Of Sociomateriality In The Context Of Critical Care. MIS Quarterly, 38(3), 895–925. https://doi.org/10.25300/misq/2014/38.3.12

Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). A Multiagent Competitive Gaming Platform to Addess Societal Challenges. MIS Quarterly, 40(2), 447–460. https://doi.org/https://doi.org/10.25300/MISQ/2016/40.2.09

Mathisen, B. M., Haro, P., Hanssen, B., Bjork, S., & Walderhaug, S. (2016). Decision Support Systems in Fisheries and Aquaculture: A systematic review. Aquaculture Engineering.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform Revolution – How Networked Markets Are Transforming the Economy–and How to Make Them Work for You. W. W. Norton & Company.

Pettigrew, A. M., Woodman, R. W., & Cameron, K. S. (2001). Studying organizational change and development: Challenges for future research. Academy of Management Journal, 44(4), 697–713.

Piplani, D., Ramesh, N., Singh, D. K., Kumar, A., Srinivasan, K., & Kumar, V. (2015). Digital Platform for Data Driven Aquaculture Farm Management. ACM International Conference Proceeding Series, 95–101. https://doi.org/10.1145/2835966.2836277

Rochet, J.-C., & Tirole, J. (2003). Two-Sided Markets. Journal of the European Economic Association, 1(4), 990–1029. https://doi.org/10.1111/j.1467-8462.2013.12020.x

Rodon, J. (2018). Platformization As Generative Entrenchment: a Case of a Public Healthcare System. Pre-ECIS 2018 Workshop on "Platformization of the Public Sector", (July).

Rodriguez, J., & Piccoli, G. (2018). Uncovering the digital "x" phenomena in the IS field: A text analysis approach. 31st Bled EConference: Digital Transformation: Meeting the Challenges, BLED 2018, 455–462. https://doi.org/10.18690/978-961-286-170-4.31

Secretan, P. A. D. (2008). Aquaculture insurance industry risk analysis processes. FAO Fisheries and Aquaculture Technical Paper, No. 519(Understanding and applying risk analysis in aquaculture), 229–245.

Shi, Z., Lee, G. M., & Whinston, A. B. (2016). Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. MIS Quarterly, 40(4), 1035–1053.

Stigebrandt, A., Aure, J., Ervik, A., & Hanson, P. K. (2004). Regulating the local environmental impact of intensive marine fish farming: III. A model for estimation of the holding capacity in the modelling-ongrowing fish farm-monitoring system. Aquaculture, 1–4, 239–261. https://doi.org/https://doi.org/10.1016/j.aquaculture.2003.11.029

Subasinghe, R. P., Curry, D., Mcgladdery, S. E., & Bartley, D. (2003). Recent Technological Innovations in Aquaculture. Review of the State of World Aquaculture. FAO Fischeries Circular, 59–74.

Syed, R., & Dhillon, G. (2015). Dynamics of data breaches in online social networks: Understanding threats to organizational information security reputation. International Conference on Information Systems: ICIS 2015, (2015), 1–17.

Tiwana, A., & Bush, A. A. (2014). Spotting lemons in platform markets: A conjoint experiment on signaling. IEEE Transactions on Engineering Management, 61(3), 393–405. https://doi.org/10.1109/TEM.2014.2311074

Vidaurre, C., Kawanabe, M., Bünau, P. von., Blankertz, B., & Müller, K. R. (2011). Toward Unsupervised Adaptation of LDA for Brain–Computer Interfaces. IEEE Transactions on Biomedical Engineering, 58(3), 587–597. https://doi.org/10.1109/TBME.2010.2093133

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the future: Writing a Literature Review. MIS Quarterly, 26(2), xiii–xxiii.

Yoo, Y., Ola, H., & Lyytinen, K. (2010). The New Organizing Logic of Digital Innovation : An Agenda for Information Systems Research. Information Systems Research, 21(4), 724–735. https://doi.org/10.1287/isre.1100.0322

Yu, R., Leung, P., & Bienfang, P. (2006). Predicting shrimp growth: Artificial neural network versus nonlinear regression models. Aquacultural Engineering, 34(1), 26–32. https://doi.org/https://doi.org/10.1016/j.aquaeng.2005.03.003