Data and Metadata. 2025; 4:572
doi: 10.56294/dm2025572

REVIEW

Data ecosystem framework proposal to implement Food Informatics systems in agri-food chains

Propuesta de un modelo de ecosistema de datos para la implementación de sistemas informáticos alimentarios en las cadenas agroalimentarias

William Alejandro Orjuela-Garzón[1,2], Henry Cárdenas-Roa[2], Daniel Bustos-Vanegas[3], Juan Manuel Andrade-Navia[4]

[1]Grupo de Investigación Bioecono, Inntegra SAS. Ibagué, Colombia.
[2]Universidad del Tolima, Faculty of Agronomic Engineering. Ibagué, Colombia.
[3]Universidad Surcolombiana, Agroindustrial Engineering Program. Neiva, Colombia.
[4]Universidad Surcolombiana, Faculty of Economics and Administration. Pitalito, Colombia.

Cite as: Orjuela-Garzón WA, Cárdenas-Roa H, Bustos-Vanegas D, Andrade-Navia JM. Data ecosystem framework proposal to implement Food Informatics systems in agri-food chains. Data and Metadata. 2025; 4:572. https://doi.org/10.56294/dm2025572

Submitted: 08-05-2024 **Revised:** 02-09-2024 **Accepted:** 06-01-2025 **Published:** 07-01-2025

Editor: Adrián Alejandro Vitón Castillo

Corresponding Author: William Alejandro Orjuela-Garzón

ABSTRACT

Agri-food chains face permanent climate change, population growth, and water and input access challenges. These impact production, processing, and marketing processes, making the capture, processing, and analysis of the data generated by each link more complex, isolated, and independent. Extracting this information for intelligent analysis to allow the optimization of agri-food chains based on data analytics is called Food Informatics. The study paradigm has given rise to the concept of data ecosystems in agri-food chains. The aim of this study is to design a data ecosystem model for the implementation of Food Informatics systems in agri-food chains. The PRISMA methodology was implemented for the identification, screening, eligibility, and inclusion of studies from the Scopus and Clarivate databases. A total of 26 records were included in the in-depth analysis, identifying two data ecosystem types: those with integrated bidirectional views that facilitate link interoperability and others of an individual nature focused on one link. The proposed integrated data ecosystem model has as its core an ETL in GCP for Data Issued in Batches with a Data Catalog and Data Mesh-type structure, which integrates a physical and a digital layer and data infrastructure for storage, processing, visualization, curation, and interaction with the user.

Keywords: Data Management; Integrated Information System; Value Chain Data; Decision Support Systems.

RESUMEN

Las cadenas agroalimentarias se enfrentan permanentemente al cambio climático, al crecimiento demográfico y a problemas de acceso al agua y a los insumos. Todo ello afecta a los procesos de producción, transformación y comercialización, lo que hace que la captura, el tratamiento y el análisis de los datos generados por cada eslabón sean más complejos, aislados e independientes. La extracción de esta información para un análisis inteligente que permita la optimización de las cadenas agroalimentarias a partir de la analítica de datos se denomina Informática Alimentaria. El paradigma de estudio ha dado lugar al concepto de ecosistemas de datos en las cadenas agroalimentarias. El objetivo de este estudio es diseñar un modelo de ecosistema de datos para la implantación de sistemas de Food Informatics en cadenas agroalimentarias. Se aplicó la metodología PRISMA para la identificación, cribado, elegibilidad e inclusión de estudios de las bases de datos Scopus y Clarivate. Se incluyeron un total de 26 registros en el análisis en profundidad, identificando dos tipos de ecosistemas de datos: aquellos con vistas bidireccionales integradas que facilitan la interoperabilidad de los enlaces y otros de carácter individual centrados en un enlace. El modelo de ecosistema de datos integrado propuesto tiene como núcleo un ETL en GCP para Datos Emitidos en Lotes con una estructura de tipo Catálogo de Datos y Malla de Datos, que integra una capa física y otra digital e infraestructura de datos para almacenamiento, procesamiento, visualización, curación e interacción con el usuario.

Palabras clave: Gestión de Datos; Sistema de Información Integrado; Datos de la Cadena de Valor; Sistemas de Apoyo a la Toma de Decisiones.

INTRODUCTION

The digitalization of the agri-food sector has been driven globally in the last 10 years by different technological advances and by the need to monitor, optimize, and improve the transparency and traceability of food reaching the final consumer. Integrating systems and applications in the agri-food chain to improve food quality, safety, traceability, and transparency have been implemented based on the use of data collected through cultivation, processing, the supply chain, and the market. This digitalization allows different processes in the value chain to be traced and recorded, improving consumer confidence in the food chain.[1]

Agri-food chains require technologies and techniques that allow them to optimize their production, transformation (processing), and consumption processes from data-centered perspectives and under a bidirectional communication approach that allows the different links in production (cultivation), processing (food production) and final consumption to generate and consult relevant information for decision-making and market adjustment.

The integration of technologies such as Artificial Intelligence, Drones, Internet of Things (IoT), Industrial Internet of Things (IIoT), Blockchain, Cloud Computing, Edge Computing, Radio Frequency Identification (RFID), Robotics-Cobots and Digital Twins have allowed the emergence and integration of data processing systems such as precision agriculture, smart agriculture, industry 4.0, food computing and food informatics. The latter is understood as the collection, preparation, analysis, and intelligent use of data from agriculture, supply chains, food processing, and consumers to extract knowledge and develop intelligent analyses that optimize agri-chains based on data analytics.[2]

These types of data capture and processing models promote the emergence of data ecosystems, which can be understood as complex socio-technical systems in which actors contribute and collaborate to capture, publish, or consume data to obtain or add value to innovate and make business decisions.[3] Similarly, most data ecosystems present a unidirectional model that does not allow feedback between users or data providers to share knowledge in both directions.

Data ecosystem models have been primarily developed under information management systems for a particular crop[4] that allows the integration of technologies such as Zigbee, NB-IoT, LoRa, Wi-Fi, Sigfox, LTE-m, 3G, 4G, 5G, Bluetooth 4.x, and IEEE 802.15.4 in activities such as cropping under controlled conditions, open field cropping, livestock breeding, aquaculture, and aquaponics.[5,6,7,8] AgroTrace[9] and DEMETER[6] stand out among the success stories in the production environment. The emergence of these ecosystems is supported by implementing technologies associated with food informatics ecosystems (table 1).

Artificial Intelligence (AI) simulates human intelligence, allowing information to be processed and learned and problems to be solved optimally. AI components are embedded in applications for climate management, residue or waste control, nutrition management, disease detection and treatment, demand projection, quality management, inventory control, consumer analysis, and food fraud detection, and they are already implemented in the Production—Processing and Marketing links.[10,11,12,13,14,15] Drones and Uncrewed Aerial Vehicles can be adapted with different technologies to monitor activities, product application, inspection, topography, and cartography. These data are reported and stored for decision-making,[11] and the main application link is production.

The Internet of Things (IoT) is a system of interconnected sensors that communicate and interact with each other, allowing the capture of information from actuators. It has the potential to be applied in different links of the agri-food chains due to its impact on sustainability, energy consumption, manufacturing costs, security, supply chain tracking, marketing, and consumer experience[16,17]. It has already been implemented in the Production—Processing and Marketing links. In this sense, the Industrial Internet of Things (IIoT) includes sensors, devices, and industrial machines interconnected to capture and share information in real-time. Its application is primarily found in the processing link because it allows the optimization of production processes based on the real-time monitoring of process variables, quality control, traceability, logistics, and inventory management[18,19]. This technology facilitates the implementation of smart factories (processing link).

Other technologies like Blockchain,[20,21,22] Cloud Computing,[23,24,25,26] Edge Computing,[27,28,29] Radio Frequency Identification (RFID),[30,31] Robotics-Cobots,[32,33,34] and Digital Twins[17,35,36] can be implemented in the Production—

Processing and Marketing links facilitating the traceability, quality verification, and certification of origin processes. They also increase food safety, brand reputation, and end-consumer satisfaction.

The development of this type of system requires internet-based solutions that can solve communication problems in agri-food chains.[38] However, cloud applications for these sectors require complete internet coverage, which is not always available in rural areas. Therefore, hybrid structures or solutions could be presented as an answer. Hence, data ecosystems are built through the collaboration and coordination of different actors, such as producers, processors, and marketers.

These data ecosystems require modern systems to meet the challenges of managing large volumes of information, especially when there are multiple sources. In this sense, Data Mesh promotes data distribution where each segment has autonomy based on compliance with previously defined standards and principles, guaranteeing agility, scalability, and quality. Data infrastructure must be oriented from the physical layers to the data layers in the storage, processing, visualization, curation, and user-interaction stages to ensure that the deployments are integrated into the physical (links) and digital (sensing technologies) constituents.

Interoperability of data across multiple links is required to achieve such infrastructure deployment. In the context of this article, interoperability is defined as the ability to seamlessly integrate data sets from multiple sources to develop cohesive system models while minimizing the need for manual editing. Interoperability promotes data sharing and reuse.[1] This type of solution is based on relationships between data management system components, including data capture, communication, the analytical process, and knowledge extraction. Other integrated models that combine data from production, transformation (processing), and marketing links (hybrid integration) have also been studied.

Aggregated functional requirements to achieve the abovementioned are presented in four categories: 1) real-time virtualization, 2) logistics connectivity, 3) logistics intelligence, and 4) configuration.[38] The Internet of Things (IoT) and Internet of Services (IoS) allow the combination of centralized and decentralized architectural proposals that improve sensor efficiency and communication. In this sense, it is necessary to study the collaborative bidirectional data ecosystem models that address the reality of agri-food chains from their limitations and opportunities and facilitate their actual implementation.

From this perspective, multiple structures and architectures of data systems can arise both from the need for a link, with its own technology and data, and from the creation of value for the different actors in the chain and, in this way, guarantee the quality and safety of food and impact the sustainable practices of the supply chain.[39] These types of structures require mechanisms that facilitate data exchange (interoperability) and improve the analysis of the same through the system integrated by the different links, where cybersecurity and trust between the parties are essential challenges to address.[40]

Data ecosystems for agri-food chains become tools designed to help decision-makers analyze complex information and make evidence-based decisions. They apply mathematical models and historical and real-time data to generate valuable insights that allow solving potential problems or anticipating challenges and opportunities, contributing to their sustainability. Decision support systems (DSS)[41] assist these tools. Accordingly, the aims of this study are:

- Map data ecosystem models at different levels for their comprehension and analysis.
- Identify the dynamics of scientific production on this topic at a global level.
- Design a proposal for a data ecosystem model for implementing Food Informatics systems in agri-food chains.

The contribution of the article focuses on the proposal of a data ecosystem model that allows communication, consultation, processing, analysis, and decision-making among the participating links from a bidirectional perspective based on data. This model integrates the perspectives of the food product producer, processor, and marketer, including consumer market data, to promote improvement processes.

The paper is structured as follows: Section 2 presents the materials and methods applied in this paper; Section 3 analyzes results at various levels, including metrics, knowledge structures, and some identified structures of data ecosystems; a proposal for an integrated data ecosystem model for agri-food chains is also presented. Finally, Section 5 includes the conclusions from this study and recommendations for future research.

METHOD

The PRISMA (Preferred Reporting Items for Systematic Reviews and Me-ta-Analyses) methodology proposed by Moher et al.[42] was followed to develop the systematic literature review process. This methodology proposes four stages: i) Identification, ii) Screening, iii) Eligibility, and iv) Inclusion. Keywords were used for the identification stage, and a search for records was carried out using two databases, Scopus and Web of Science (WoS). All available collections were used for the search, including Core Collection, which covers the Sciences, Social Sciences, and Arts and Humanities indexes, proceedings of both Sciences and Social Sciences and Humanities, and complementary databases (Medline, Scielo, and Korean Citation Index, BIOSIS Previews, and FSTA – The Food Science Resource). These were explored in August 2024, using structured search equations based on Boolean operators and field codes (title – abstract – keywords) for the last 10 years. Table 1 shows the process of constructing the search equations for each analysis link.

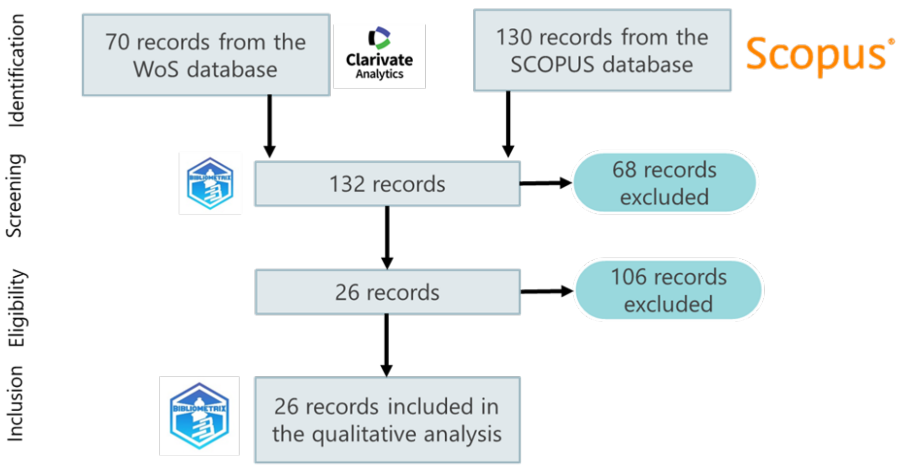
Table 1. Query algorithms and the number of results of each database used

Database	Query algorithms	# of results
Scopus	((TITLE-ABS-KEY (agro-industria)) OR (TITLE-ABS-KEY (agro-industry)) OR (TITLE-ABS-KEY (agro-industrial)) OR (TITLE-ABS-KEY (agro-industry)) OR (TITLE-ABS-KEY (agribusiness)) OR (TITLE-ABS-KEY (“agri-food”)) OR (TITLE-ABS-KEY (agroaliment*)) OR (TITLE-ABS-KEY (“agri-food sector”)) OR (TITLE-ABS-KEY (agro-food)) OR (TITLE-ABS-KEY (“food chain”)) OR (TITLE-ABS-KEY (“Food System*”))) AND (TITLE-ABS-KEY (“data sharing”) OR TITLE-ABS-KEY (“FAIR principles”) OR TITLE-ABS-KEY (“Data management”) OR TITLE-ABS-KEY (“Data ownership”) OR TITLE-ABS-KEY (“Domain analysis”) OR TITLE-ABS-KEY (“Feature model”) OR TITLE-ABS-KEY (“Reference architecture”) OR TITLE-ABS-KEY (“Data Ecosystem”)) AND PUBYEAR > 2013 AND PUBYEAR < 2025	130
Web of Science	((TI=agro-industria OR AB=agro-industria OR AK=agro-industria) OR (TI=agroindustry OR AB=agroindustry OR AK=agroindustry) OR (TI=agro-industrial OR AB=agro-industrial OR AK=agro-industrial) OR (TI=agro-industry OR AB=agro-industry OR AK=agro-industry) OR (TI=agribusiness OR AB=agribusiness OR AK=agribusiness) OR (TI=”agri-food” OR AB=”agri-food” OR AK=”agri-food”) OR (TI=agroaliment* OR AB=agroaliment* OR AK=agroaliment*) OR (TI=”agri-food sector” OR AB=”agri-food sector” OR AK=”agri-food sector”) OR (TI=agro-food OR AB=agro-food OR AK=agro-food) OR (TI=”food chain” OR AB=”food chain” OR AK=”food chain”) OR (TI=”Food System*” OR AB=”Food System*” OR AK=”Food System*”)) AND ((TI=”data sharing” OR AB=”data sharing” OR AK=”data sharing”) OR (TI=”FAIR principles” OR AB=”FAIR principles” OR AK=”FAIR principles”) OR (TI=”Data management” OR AB=”Data management” OR AK=”Data management”) OR (TI=”Data ownership” OR AB=”Data ownership” OR AK=”Data ownership”) OR (TI=”Domain analysis” OR AB=”Domain analysis” OR AK=”Domain analysis”) OR (TI=”Feature model” OR AB=”Feature model” OR AK=”Feature model”) OR (TI=”Reference architecture” OR AB=”Reference architecture” OR AK=”Reference architecture”) OR (TI=”Data Ecosystem” OR AB=”Data Ecosystem” OR AK=”Data Ecosystem”)) AND PY=(2014-2024)	70

A total of 200 records were obtained for the identification stage, extracted from the two databases consulted for the second stage, i.e., screening. Duplicate records were eliminated using the method described by Caputo and Kargina[43] in the Bibliometrix 4.1 software,[44] yielding 132 unique records. For the eligibility stage, two exclusion criteria were established:

- Articles focused on the agri-food chain (production – processing – marketing).
- Articles that clearly show the technological structure used for data acquisition, storage, analysis, and visualization in graphic or textual form.

Under these criteria, 106 documents were excluded after the in-depth review. Finally, 26 papers were selected for the qualitative and quantitative analysis at the inclusion stage. Figure 1 below graphically shows the procedure for the information elements preferred for systematic reviews and meta-analysis.



Source: adapted from Moher et al.[42]
Figure 1. PRISMA methodology

RESULTS

For the analysis of the results, the Bibliometrix 4.1 software developed by Aria et al.[44] was used, which facilitates the quantitative analysis of bibliographic records related to a specific area of knowledge based on different bibliometric techniques and supported by the R programming language. In general terms, 26 documents were obtained with an inter-annual growth rate of 14,8 % and an international co-authorship of 26,92 %. In addition, these documents have been published in more than 23 different journals by more than 129 different authors. The analysis was conducted using two levels of metrics and one level of knowledge structures. The first relates mainly to annual scientific production, primary sources, authors, and documents. The second is oriented toward identifying conceptual, intellectual, and social structures of the subject studied.

Metrics level

The scientific production related to data ecosystem models for the implementation of Food Informatics systems in agri-food chains has shown a constant interest in the last 10 years, with a total of 26 documents and a positive evolution going from one record in 2014 to four in 2024 and an average interannual growth rate of 14,87 %. The scientific production peak for the knowledge area occurs in 2021 with six documents. Although 2024 presents a partial indexing and the search was carried out in August, an increase in publications on the subject is evident (see figure 2).

The international journals that publish research related to the search topic are Computers and Electronics in Agriculture, International Journal on Food System Dynamics, and Sustainability, all with two publications. A thorough review shows that this new knowledge base related to the topic of Food Informatics data ecosystem models is supported by high-impact journals such as Computers and Electronics in Agriculture, a journal with a long history in the areas of crop science, applied computer science, among others, and with an H index of 168. Likewise, the journal Agricultural Systems appears oriented to analyzing interactions between the different hierarchical levels of agricultural systems and has an H index of 134.

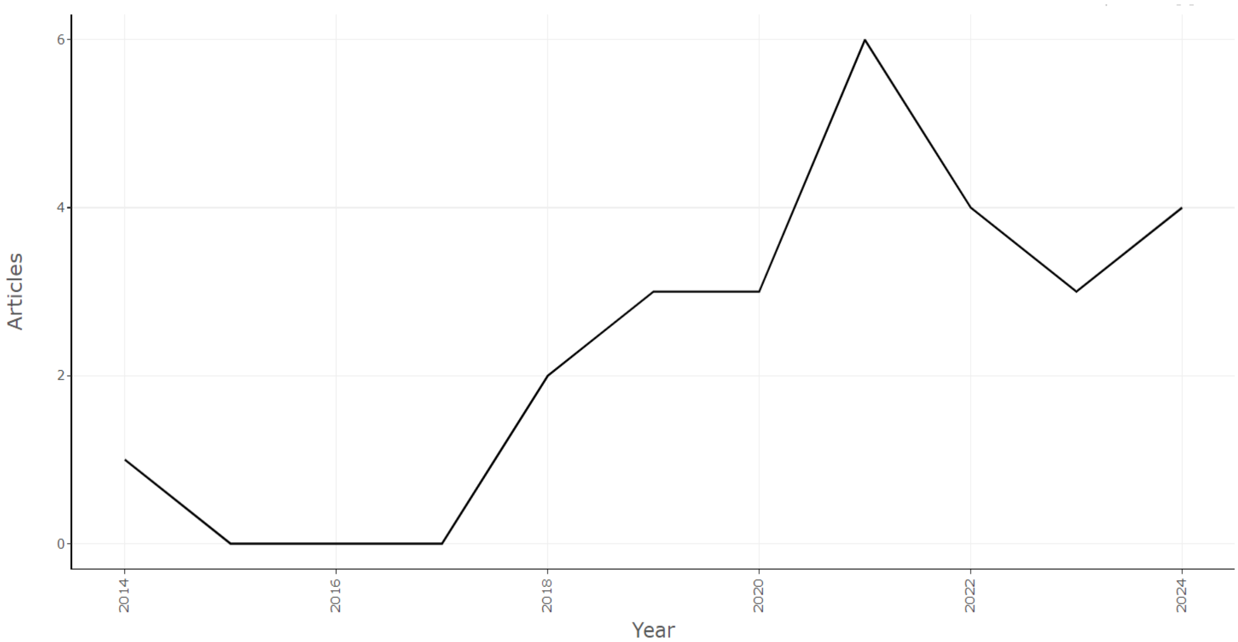


Figure 2. Scientific production (2014-2024)

Among the most cited articles in the corpus is “A Review of Edge Computing Reference Architectures and a New Global Edge Proposal,” with more than 116 citations, published by Inés Sittón-Candanedo from the Bisite research group at the University of Salamanca in Spain. This article reviews Edge Computing technology and proposes a modular reference architecture to address the complexity of the implemented solutions. The contribution of Blockchain technologies in terms of security and privacy is highlighted. Finally, a use case is described to validate the proposed architecture in a smart agro-industry scenario. Another article of relevance to the knowledge base in the corpus is “A Life Cycle Framework of Green IoT-Based Agriculture and its Finance, Operation, and Management Issues,” with nearly 100 citations, published by Junhu Ruan from Northwest A&F University in China. This article focuses on the challenges agricultural producers face in implementing IoT technologies to increase crop yield, especially when these technologies require complex infrastructure models.

Among the leading institutions hosting researchers in the field of data ecosystems for the implementation of food informatics systems, Wageningen University in the Netherlands stands out with eight publications, as well as the National Institute for Research in Agriculture, Food and Environment of France. Regarding Wageningen University, its research in data ecosystems is oriented toward the definition of principles for collecting, sharing, and using data in the agri-food sector under elements of privacy and trust, as well as codes of conduct.

Among the leading researchers in data ecosystems in the agri-food chains area, Bedir Tekinerdogan from the Information Technology Group at Wageningen University stands out. His publications focus on the construction or definition of reference models or architectures to establish the requirements for collaboration between diverse institutions seeking to share data. The countries that concentrate the publications in the study area are the Netherlands, France, and China, with 11, 8, and 6 publications, respectively.

Level of knowledge structures

The methodology proposed by Cobo et al.[45] was followed to analyze knowledge structures. This methodology establishes two arguments: density and centrality. In this case, the number of documents, the number of citations, and the average number of citations were considered to identify emerging, basic, niche, or declining topics.

From the perspective of the conceptual structure of the analyzed records, it is evident that there are four cores or dimensions of topics that guide the research trends in this field according to the density (degree of development of the topic) concerning the centrality (degree of relevance) as indicated in figure 3.

The first core highlights emerging themes directly related to ontological models or reference architectures for data ecosystems in the agri-food sector using technologies such as IoT and Blockchain. A second core includes driving themes, that is, those with high relevance and degree of development in the corpus analyzed and are directly related to interoperability and data sharing schemes, critical topics in the analyzed research because they are considered a barrier to the implementation of food informatics technologies in agri-food chains. A third core is related to niche or highly specialized topics such as decision support systems based on food informatics systems and the operational sustainability of this type of technology from a business development perspective due to investment, maintenance, and operation costs for its deployment. Finally, there is a fourth core with so-called basic topics, i.e., those with a degree of relevance; their level of development is lower, or the literature is already saturated. In this sense, implementing Big Data in agriculture is a widely addressed topic. Developing and implementing food informatics systems in agri-food chains is crucial, especially in the data analytics component, because the different links generate data that must be integrated and analyzed.

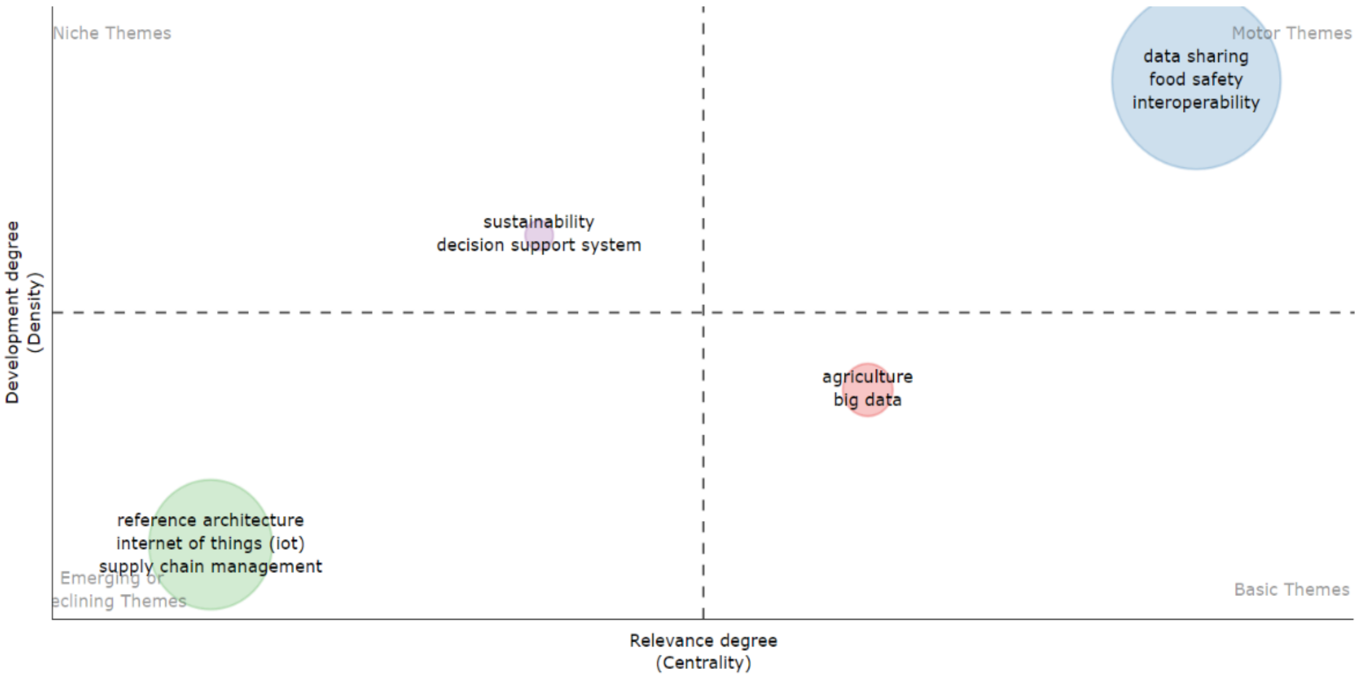


Figure 3. Conceptual structure – Thematic Map

Regarding the collaboration networks between authors, two nodes stand out. The first node in red with a central core led by Bedir Tekinerdogan from Wageningen University and Cor Verdouw from the company Mprise Agriware Mprise B.V. (More information available at https://www.mprise-agriware.com/) that develops and markets software for the operational and administrative management and data analysis and decision making in agri-food chains (see figure 4). Krijn Poppe leads the second node in blue from Wageningen University and Roeland van Dijk from the company Soops Group. This company group developed an innovative solution in the fields of data storage, analysis, and management of workflows and processes under the name ARTIS. Over time, ARTIS has been developed and transformed into SITRA. This new system is more advanced and integrates data from various open and closed sources. SITRA can collect and protect these data, enrich them, and make predictions. These capabilities are essential to support administrative processes, regulatory compliance, and decision-making, making SITRA an indispensable tool for efficient management and strategic planning for agri-food chains.

In this sense, the authors mentioned reflect a strong institutional interaction where the main node is concentrated in Wageningen University, which is linked to the University of Qatar and Griffith University located in Australia, as well as to the company Mprise Agriware Mprise B.V. On the other hand, the country that concentrates the publication dynamics is the Netherlands (green node), which is directly linked to Australia and creates interactions with another blue node led by Spain through a connection with Irish researchers. A red node stands out, mainly linking Asian countries such as China, Japan, and Malaysia with Spain for knowledge generation (see figure 5). Finally, an isolated node between researchers from countries such as Italy and Switzerland can be observed.

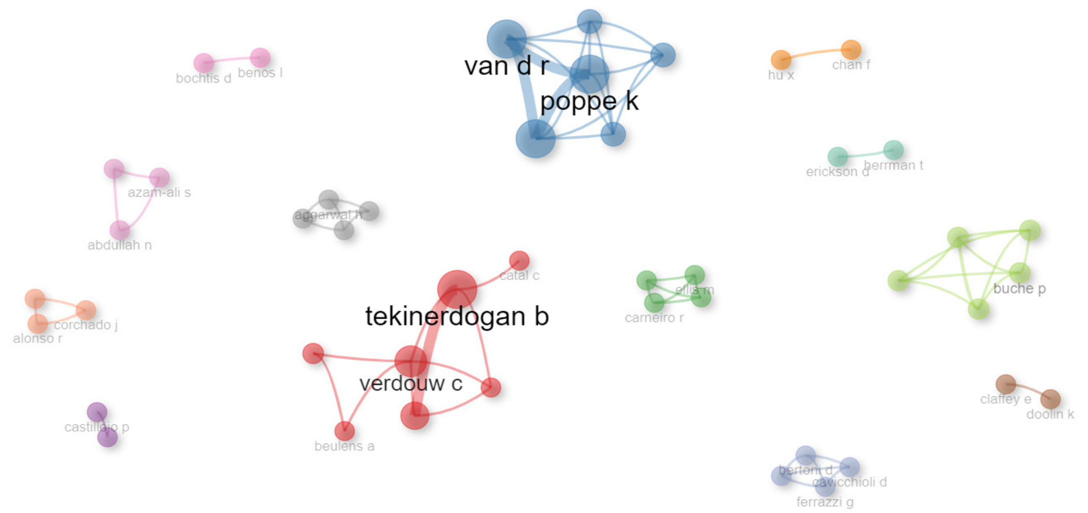


Figure 4. Authors collaboration network

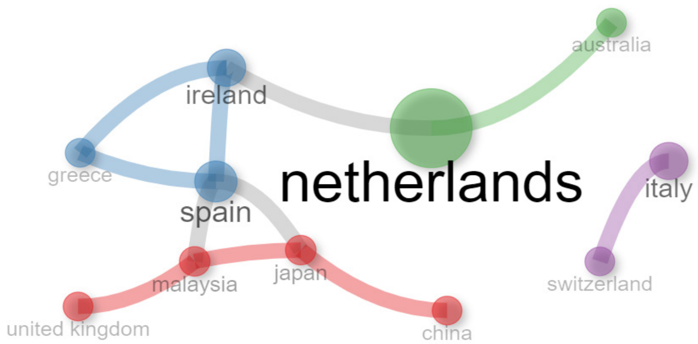


Figure 5. Country collaboration network

Carrying out an in-depth analysis of the corpus and considering inclusion criteria such as i) focusing on the agri-food chain (production – processing – marketing) and ii) clearly presenting the technological structure used for data acquisition, storage, analysis, and visualization in graphic or textual form, two main categories were identified: 1) articles focused on the implementation and/or development of data ecosystems in the production link, and 2) articles transversal to the entire agri-food chain. Table 2 shows a summary for each category, including the title, abstract, and reference of the article.

Data ecosystems in the production chain

For their part, the articles referring to the production link mainly explore precision agriculture activities to increase yields, trying to overcome challenges such as the enormous initial investment and the difficulty of acceptance of the technologies by producers by implementing sensors and communication technologies such as Zigbee, NB-IoT, LoRa, Wi-Fi, Sigfox, LTE-m, 3G, 4G, 5G, Bluetooth 4.x, and IEEE 802.15.4, in activities such as cultivation under controlled conditions, open field cultivation, livestock breeding, aquaculture and aquaponics.[5,6,7]

Table 2. Articles associated with Data Ecosystems in the production link

Title	Summary	Reference
A life cycle framework of green IoT-based agriculture and its finance operation and management issues	Addresses the challenges and opportunities in implementing IoT systems in agriculture, highlighting their potential to improve agricultural yields through precision agriculture, expanding the focus of IoT implementation from the growth cycle to the entire life cycle of agricultural products, and considering energy efficiency.	(5)
A review of edge computing reference architectures and a new global edge proposal	Reviews edge computing technology and its reference architectures, highlighting their ability to minimize latency, improve privacy, and reduce bandwidth costs in IoT scenarios.	(46)
Building an interoperable space for smart agriculture	DEMETER presents a comprehensive solution to meet the challenges of digital transformation in agriculture, especially regarding interoperability and data management. DEMETER integrates heterogeneous hardware and software resources to facilitate data and knowledge exchange in the agri-food chain.	(6)
Cropinfra - An Internet-based service infrastructure to support crop production in future farms	The Cropinfra platform was developed and evaluated. It is a multi-level service framework designed to improve precision agriculture through sensors, data collection and control, storage, and external services such as forecasting.	(7)
Design of a system for information transfer to reduce administrative burdens in the agri-food sector	It presents the design of the SITRA system, which combines data from various sources and allows farmers to control who can access their data, thereby reducing administrative burdens.	(8)
Enablers to digitalization in agriculture: A case study from Italian field crop farms in the Po River valley with insights for policy targeting	The drivers and barriers to adopting digitalization in field crop farms in the Po Valley are explored, identifying key factors such as digital skills, data management practices, and interoperability. It proposes that policies should focus not only on the acquisition of equipment but also on the development of the human capital of farmers.	(4)
Integrating IOTAs tangle with the Internet of Things for sustainable agriculture: A proof-of-concept study on rice cultivation	The study analyses how integrating IoT technologies with IOTA Tangle –a distributed ledger technology (DLT)– can improve sustainable agricultural practices, using rice cultivation as an example. Smart irrigation systems were implemented, significantly reducing water use and nitrogen and methane emissions while improving resource management through real-time monitoring.	(47)
Reference architecture design for developing data management systems in smart farming	Proposes a reference architecture for data management in smart agriculture developed through domain analysis and architectural modeling.	(48)
Securing data in life sciences - A plant food (Edamame) systems case study	It addresses challenges in designing specialty crops, such as edamame, that improve sustainability and yields, highlighting the importance of integrating data on nutritional content, sensory profile, and consumer acceptability. The need for improved communication and collaboration between the agricultural and food sectors to address cyber-biosecurity risks and improve data sharing is emphasized.	(40)
Underutilised crops database for supporting agricultural diversification	It presents the creation of a globally accessible database to store information on underutilized crops, an important step towards improving agrobiodiversity in digital agriculture. A robust relational data model was adopted, and an open-access web user interface was developed to facilitate data management and access.	(49)

Edge computing is presented as an innovative solution that brings local machines as close to sensors as possible. This minimizes latency, improves privacy, and optimizes bandwidth usage in IoT scenarios.[46]

The scientific literature presents other success stories in the production environment, such as AgroTrace and DEMETER, architectures that have allowed the optimization of information flows and recognize the importance of developing systems that will enable interoperability between heterogeneous production systems and permit addressing multiple scenarios of particular agricultural sectors seeking solutions that allow overcoming limitations in terms of the possibility of generalizing and applying the architectures (see figure 6).[6]

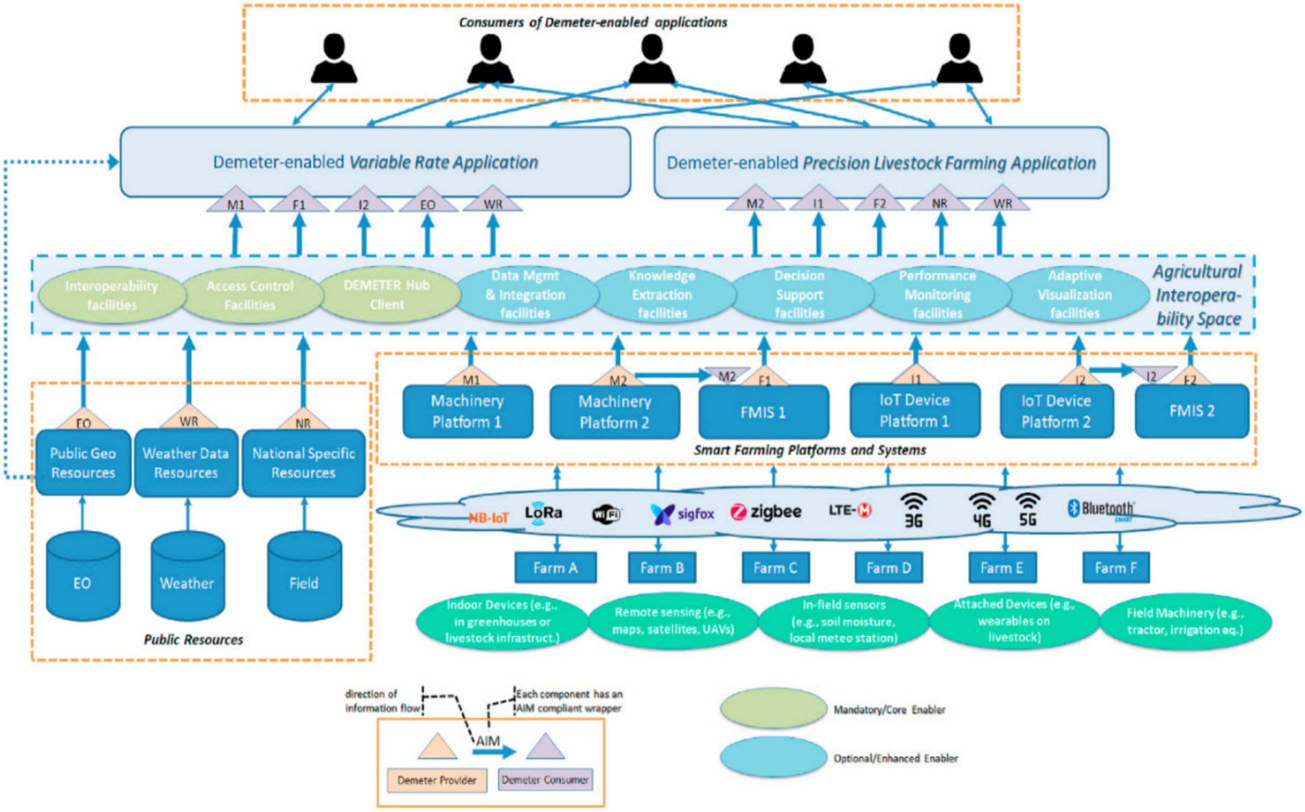


Figure 6. High-level vision of DEMETER[6]

Figure 7 presents a diagram that reflects the relationship of components proposed in most documents regarding the production link. This relationship shows important similarities with those proposed transversally to the complete cycle.[48] However, the flow of information and knowledge is unidirectional.

Among the uses reported are constructing government interfaces and disseminating data for decision-making concerning the use of crops to conserve agrobiodiversity and improve yields in the production link.[49] Table 3 shows the scientific literature identified in the corpus and associated with the production link.

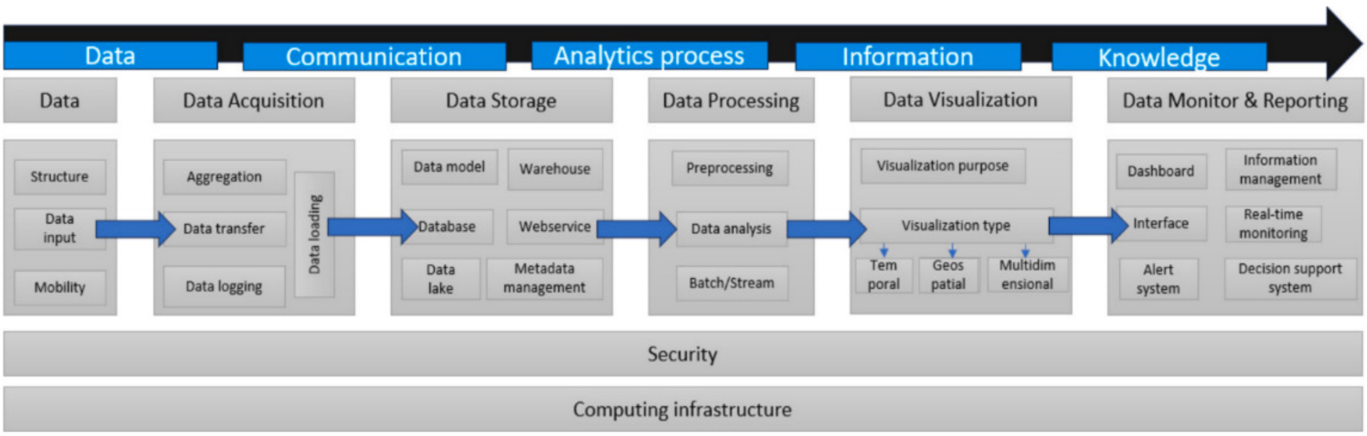


Figure 7. List of components of a data management system for smart farming[48]

Data ecosystems integrated into the entire agri-food chain

The reviewed articles propose a hybrid integration that allows stakeholders throughout the chain to access information of interest during their logistics processes. The core of the architecture is a modular cloud-based smart logistics system that, in turn, interacts with local systems that control the actuators and store sensory and logistics information of interest to synchronize it by event, thus optimizing the use of the network and the robustness of information storage.[38]

Data integration addresses two key concepts associated with this architecture: "Big Data" and "Decision Support Systems" (DSS). This allows organizations to make informed, data-oriented decisions that help them identify and solve problems.[50]

Figure 8 shows the proposed flow of information from the sources of information to decision-making.

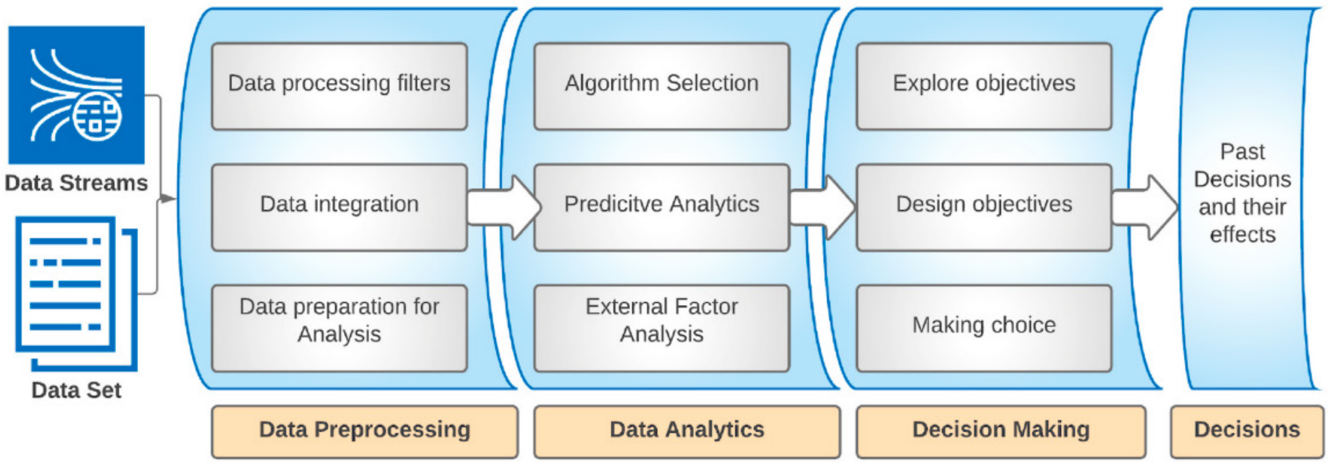


Figure 8. Big Data analytics structure for Decision Support Systems[50]

As a support to the architecture and to facilitate access to information throughout the chain, a search engine is presented with a model of integration, organization, and reuse of data related to the quality, safety, traceability, and authenticity of food in supply chains that allow information to be made available for all interested parties to access, presenting as a prototype AgroTrace. It is an open-access traceability system that incorporates IoT technologies and allows monitoring of product quality throughout the entire supply chain, from the field to the consumer, with extension to residue or waste management, promoting a circular economy by comparing different ranges in the chain such as field-retailer or field-consumer and, in turn, identifying the leading information transfer technologies such as RFID, BLE, GSM, NFC, and LoRaWAN (see figure 9).[9,51]

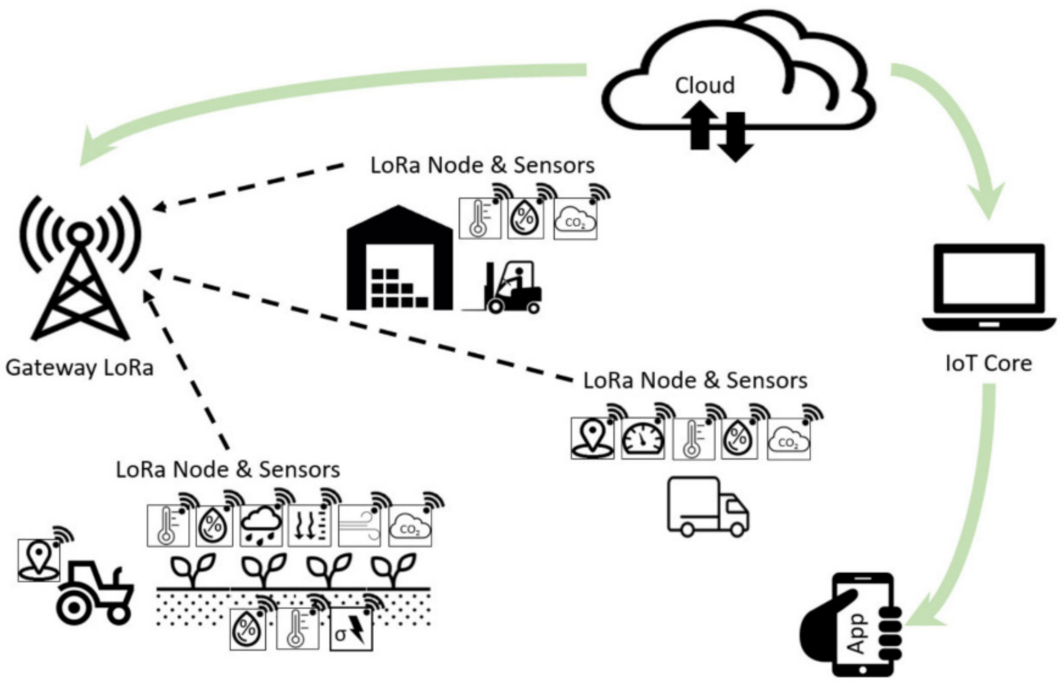


Figure 9. Internet of Things (IoT) implementation scheme using LoRa[9]

Beyond technological implementations, a design framework is also defined to help align business and IT processes in collaborating organizations using a common reference architecture.[52] Table 3 presents the literature associated with data ecosystems that integrates the different links in the agri-food chain and serves as a reference for the proposal of the data ecosystem model for implementing Food Informatics systems in agri-food chains.

Table 3. Articles associated with data ecosystems integrated into the entire agri-food chain

Title	Summary	Reference
A reference architecture for IoT-based logistic information systems in agri-food supply chains	It proposes a reference architecture that combines IoT and cloud computing to improve remote monitoring of deliveries and product conditions in agri-food supply chains, facilitating customized and affordable solutions.	(38)
A search engine concept to improve food traceability and transparency: Preliminary results	The model focuses on integrating, organizing, and reusing food quality, safety, traceability, and authenticity data in specific supply chains, such as olive oil, milk, and fishery products. This tool allows food chain actors to visualize and reuse available information quickly.	(51)
BITA*: Business IT alignment framework of multiple collaborating organisations	It presents a design framework that helps align business and IT processes in collaborating organizations using a common reference architecture. This framework was developed and validated in a real case study, especially in the agri-food sector, and addresses the challenges of comparing models from different organizations.	(52)
Bridging the gaps in traceability systems for fresh produce supply chains: Overview and development of an integrated IoT-based system	Proposes AgroTrace, an open-access traceability system that incorporates IoT technologies and allows product quality monitoring throughout the entire supply chain, from the field to the consumer, with extension to residue or waste management, promoting a circular economy.	(9)
State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change	Provides a framework for researchers and stakeholders in the agri-food sector to understand and use Big Data and DSS to improve food safety and decision-making.	(50)

In conclusion, cloud infrastructure is identified as the core for articulating information from different data sources (sensors, business management software, production management software, and analog data) with web interfaces that allow each of the actors in the production chain to record and consult information, as well as feed estimation and analytical models to strengthen these organizations, transforming them into data-oriented organizations.

Proposal for an integrated data ecosystem model for agri-food chains

An integrated model of data ecosystems for agri-food chains is proposed based on the elements identified in the literature review. Through different interfaces, this model allows capturing and consulting information from each of the actors in the chain, achieving a transformation in the participating organizations from a data analysis and decision-making perspective that optimizes the aggregation of market-oriented value in each link in the agri-food chain (see figure 10).

The integrated data ecosystem model has as its core the Extract, Transform, and Load (ETL) in Google Cloud Platform (GCP) for Batch Issued Data with Data Catalog and Data Mesh type structure in which the following conditions must be met:

- Data domains: data is organized and managed in specific domains (e.g., Sales, Marketing, Finance), each with its own responsibilities for the quality and processing of its data.
- Product ownership: each domain treats its data as a product, with a dedicated team responsible for the development, maintenance, and evolution of its data.
- Interoperability: despite the autonomy of each domain, data is designed to be interoperable, allowing for easy integration and consumption by other domains.
- Federated Governance: data governance, security, and quality policies are defined at a central level but enforced at the domain level, ensuring consistency and compliance without sacrificing flexibility.[53]

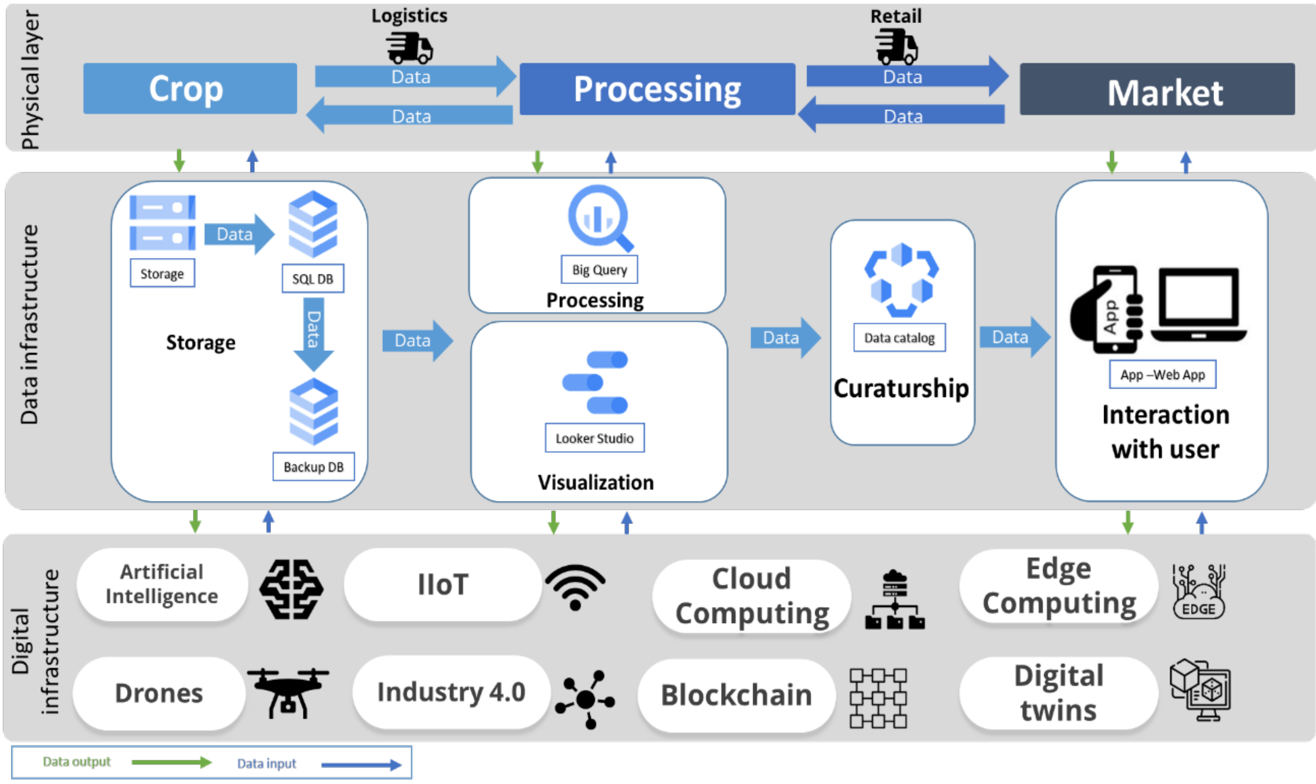


Figure 10. Integrated cloud-centered data ecosystem model for agri-food chains under the Food Informatics approach

With this architecture in mind, the proposal is focused mainly on the following phases and products:

1. Data Ingestion:
    - Cloud Storage: batch data are stored in Google Cloud Storage (GCS). A structure organized by data domains (e.g., sales, marketing, operations) is recommended to reflect a data mesh architecture. Each domain has its own bucket or set of buckets.

2. Data Catalog:
    - Metadata: Uses Google Cloud Data Catalog to index or catalog and organize the metadata for all data stored in GCS and BigQuery.
    - Documentation: Each data domain should have its own metadata descriptors, including data source, transformations, quality, and access policies.
    - Search: Implements governance policies to facilitate data search and access within the organization, ensuring that teams can efficiently discover and use data from other domains.

3. Preprocessing:
    - Cloud Dataflow: it is used to handle data preprocessing by domain. Each domain team is responsible for developing and maintaining its own data processing pipelines.
    - Standardization: Although domain teams are autonomous, common guidelines (transformation and validation patterns) are established to ensure data is interoperable between domains.

4. Data Loading:
    - BigQuery: the processed data is loaded into BigQuery, where each domain has its own dataset. Tables within each dataset should be partitioned by relevant criteria (date, region, among others) to optimize performance and scalability.
    - Interoperability: defines clear data-sharing policies between domains, allowing datasets from one domain to be accessible to other domains under established governance rules.

This design guarantees efficient data management. Each domain team is autonomous and responsible for its own data while maintaining consistency and governance at the organizational level.

CONCLUSIONS

The cloud is identified as the core for articulating information from various sources, such as sensors, business management software, and production management systems. Web interfaces allow different actors in the agri-food chain to record and consult information and, in turn, feed predictive and analytical models to transform organizations into data-oriented entities.

A Data Mesh structure is recommended to achieve an infrastructure proposal that articulates available information sources. In this structure, data is organized into specific domains (e.g., production, transformation (processing), and marketing), each responsible for the quality and processing of its data. This approach allows decentralized management, maintaining interoperability between the different domains of the organization.

This model guarantees efficient data management. Each domain has autonomy and responsibility over its own data while maintaining coherence and governance at the organizational level. The system is scalable and can adapt to the changing needs of data-oriented agri-food chains.

The proposed framework is particularly effective for integrating diverse data sources in the production, processing, and distribution phases, improving decision-making, efficiency, and traceability in agri-food ecosystems.

Finally, it is essential to highlight success stories in the production environment, such as AgroTrace and DEMETER, architectures that have allowed the optimization of information flows and recognize the importance of developing systems that permit interoperability between heterogeneous production systems and countenance addressing multiple scenarios of particular agricultural sectors.

Hybrid models of capture, processing, analysis, and decision-making based on data for agri-food chains under food informatics systems and with specific domains are identified as potential data ecosystems for agri-food chains due to the technological, financial, and human heterogeneity present in chain links.[37] This type of open data ecosystem is characterized by presenting multiple interdependencies at the levels and dimensions of socio-technical systems;[3] therefore, the interaction between the parties is complex and must be harmonized before implementation to define the data structures so the security and privacy of the data are established, and finally, to estimate the associated costs that this will have.

Decision-making processes arising from these ecosystems must consider systemic effects that flow forward or backward to other chain members. In other words, if the processing and marketing links define the development of a particular product, these decisions will likely affect the cultivation processes to ensure that the raw materials meet specific demand-side requirements.

ACKNOWLEDGMENTS

This study has been made possible thanks to the generous funding of the Ministry of Science, Technology and Innovation MINCIENCIAS, to whom we express our most sincere gratitude. We also acknowledge and thank the invaluable contribution of all the people who, directly or indirectly, have participated in this research.

BIBLIOGRAPHIC REFERENCES

1. Fritter M, Lawrence R, Marcolin B, Pelletier NA. Survey of Life Cycle Inventory Database Implementations and Architectures, and Recommendations for New Database Initiatives. Int. J. Life Cycle Assess. 2020; 25: 1522–1531.

2. Krupitzer C, Stein A. Food Informatics—Review of the Current State-of-the-Art, Revised Definition, and Classification into the Research Landscape. Foods. 2021; 10: 2889.

3. Marcelo MI, Barros Lima G de F, Farias Lóscio B. Investigations into Data Ecosystems: A Systematic Mapping Study. Springer London. 2019; 61.

4. Giorgio A, Penate Lopez LP, Bertoni D, Cavicchioli D, Ferrazzi G. Enablers to Digitalization in Agriculture: A Case Study from Italian Field Crop Farms in the Po River Valley, with Insights for Policy Targeting. Agriculture. 2024; 14: 1074.

5. Ruan J, Wang Y, Chan FTS, Hu X, Zhao M, Zhu F, Shi B, Shi Y, Lin F. A Life Cycle Framework of Green IoT-Based Agriculture and Its Finance, Operation, and Management Issues. IEEE Commun. Mag. 2019; 57: 90–96.

6. Roussaki I, Doolin K, Skarmeta A, Routis G, Lopez-Morales JA, Claffey E, Mora M, Martinez JA. Building an Interoperable Space for Smart Agriculture. Digit. Commun. Networks. 2023; 9: 183–193.

7. Pesonen LA, Teye FKW, Ronkainen AK, Koistinen MO, Kaivosoja JJ, Suomi PF, Linkolehto RO. Cropinfra - An Internet-Based Service Infrastructure to Support Crop Production in Future Farms. Biosyst. Eng. 2014; 120: 92–101.

8. Poppe K, Vrolijk H, van Dijk R. Design of a System for Information Transfer to Reduce Administrative Burdens in the Agri-food Sector. Int. J. Food Syst. Dyn. 2021; 12: 301–313.

9. Tagarakis AC, Benos L, Kateris D, Tsotsolas N, Bochtis D. Bridging the Gaps in Traceability Systems for Fresh Produce Supply Chains: Overview and Development of an Integrated Iot-based System. Appl. Sci. 2021; 11(6): 7596.

10. Bačiulienė V, Bilan Y, Navickas V, Civín L. The Aspects of Artificial Intelligence in Different Phases of the Food Value and Supply Chain. Foods. 2023; 12: 1654.

11. Spanaki K, Karafili E, Sivarajah U, Despoudi S, Irani Z. Artificial Intelligence and Food Security: Swarm Intelligence of AgriTech Drones for Smart AgriFood Operations. Prod. Plan. Control. 2022; 33: 1498–1516.

12. Talaviya T, Shah D, Patel N, Yagnik H, Shah M. Implementation of Artificial Intelligence in Agriculture for Optimisation of Irrigation and Application of Pesticides and Herbicides. Artif. Intell. Agric. 2020; 4: 58–73.

13. Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A. IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry. IEEE Internet Things J. 2022; 9: 6305–6324.

14. Al-Sarayreh M, Gomes Reis M, Carr A, Reis MM. dos Inverse Design and AI/Deep Generative Networks in Food Design: A Comprehensive Review. Trends Food Sci. Technol. 2023; 138: 215–228.

15. Buyuktepe O, Catal C, Kar G, Bouzembrak Y, Marvin H, Gavai A. Food Fraud Detection Using Explainable Artificial Intelligence. Expert Syst. 2023; 13387.

16. Dadhaneeya H, Nema PK, Arora VK. Internet of Things in Food Processing and Its Potential in Industry 4.0 Era: A Review. Trends Food Sci. Technol. 2023; 139: 104109.

17. Wu B, Gao B, Xu W, Wang H, Yi R P. Sustainable Food Smart Manufacturing Technology. Inf. Process. Manag. 2022; 59: 1–11.

18. Grover J. Industrial IoT and Its Applications. In: IoT for Sustainable Smart Cities and Society. 2022; 107–124.

19. Redchuk A, Walas Mateo F, Pascal G, Tornillo JE. Adoption Case of IIoT and Machine Learning to Improve Energy Con-sumption at a Process Manufacturing Firm, under Industry 5.0 Model. Big Data Cogn. Comput. 2023; 7: 42.

20. Treiblmaier H. The Impact of the Blockchain on the Supply Chain: A Theory-Based Research Framework and a Call for Action. Supply Chain Manag. An Int. J. 2018; 23: 545–559.

21. Treiblmaier H, Garaus M. Using Blockchain to Signal Quality in the Food Supply Chain: The Impact on Consumer Purchase Intentions and the Moderating Effect of Brand Familiarity. Int. J. Inf. Manage. 2023; 68: 102514.

22. Mohammed A, Potdar V, Quaddus M, Hui W. Blockchain Adoption in Food Supply Chains: A Systematic Literature Review on Enablers, Benefits, and Barriers. IEEE Access 2023; 11: 14236–14255.

23. Rajak ARA. Emerging Technological Methods for Effective Farming by Cloud Computing and IoT. Emerg. Sci. J. 2022; 6: 1017–1031.

24. Duan P, Wang W, Zhang W, Gong F, Zhang P, Rao Y. Food Image Recognition Using Pervasive Cloud Computing. In Proceedings of the 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing; IEEE. 2013; 1631–1637.

25. Guo Y, Chen Y, Wei W, Li H. Research on Food Recommendation Method Based on Knowledge Graph. In; 2023; pp. 521–533.

26. Gilal NU, Al-Thelaya K, Al-Saeed JK, Abdallah M, Schneider J, She J, Awan JH, Agus M. Evaluating Machine Learning Technologies for Food Computing from a Data Set Perspective. Multimed. Tools Appl. 2023; 83: 32041–32068.

27. Liu C, Cao Y, Luo Y, Chen G, Vokkarane V, Yunsheng M, Chen S, Hou P. A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure. IEEE Trans. Serv. Comput. 2018; 11: 249–261.

28. Akhtar MN, Shaikh AJ, Khan A, Awais H, Bakar EA, Othman AR. Smart Sensing with Edge Computing in Precision Agriculture for Soil Assessment and Heavy Metal Monitoring: A Review. Agricultur. 2021; 11: 475.

29. Gai K, Fang Z, Wang R, Zhu L, Jiang P, Choo KKR. Edge Computing and Lightning Network Empowered Secure Food Supply Management. IEEE Internet Things J. 2022; 9: 14247–14259.

30. Soltani Firouz M, Mohi-Alden K, Omid M. A Critical Review on Intelligent and Active Packaging in the Food Industry: Research and Development. Food Res. Int. 2021; 141: 110113.

31. Ghaani M, Cozzolino CA, Castelli G, Farris S. An Overview of the Intelligent Packaging Technologies in the Food Sector. Trends Food Sci. Technol. 2016; 51: 1–11.

32. Carolan M. Automated Agri-food Futures: Robotics, Labor and the Distributive Politics of Digital Agriculture. J. Peasant Stud. 2020; 47: 184–207.

33. Iqbal J, Khan ZH, Khalid A. Prospects of Robotics in Food Industry. Food Sci. Technol. 2017; 37: 159–165.

34. Duong LNK, Al-Fadhli M, Jagtap S, Bader F, Martindale W, Swainson M, Paoli A. A Review of Robotics and Autonomous Systems in the Food Industry: From the Supply Chains Perspective. Trends Food Sci. Technol. 2020; 106: 355–364.

35. Hassoun A, Aït-Kaddour A, Abu-Mahfouz AM, Rathod NB, Bader F, Barba FJ, Biancolillo A, Cropotova J, Galanakis CM, Jambrak AR, et al. The Fourth Industrial Revolution in the Food Industry—Part I: Industry 4.0 Technologies. Crit. Rev. Food Sci. Nutr. 2023; 63: 6547–6563.

36. Li Y, Su DA, Mardani A. Digital Twins and Blockchain Technology in the Industrial Internet of Things (IIoT) Using an Extended Decision Support System Model: Industry 4.0 Barriers Perspective. Technol. Forecast. Soc. Change. 2023; 195: 122794.

37. Orjuela-Garzon WA, Sandoval-Aldana A, Mendez-Arteaga JJ. Systematic Literature Review of Barriers and Enablers to Im-plementing Food Informatics Technologies: Unlocking Agri-Food Chain Innovation. Foods. 2024; 13: 3349.

38. Verdouw CN, Robbemond RM, Verwaart T, Wolfert J, Beulens AJM. A Reference Architecture for IoT-Based Logistic Information Systems in Agri-Food Supply Chains. Enterp. Inf. Syst. 2018; 12: 755–779.

39. Saurabh S, Dey K. Blockchain Technology Adoption, Architecture, and Sustainable Agri-Food Supply Chains. J. Clean. Prod. 2021; 284: 124731.

40. Duncan SE, Zhang B, Thomason W, Ellis M, Meng N, Stamper M, Carneiro R, Drape T. Securing Data in Life Sciences—A Plant Food (Edamame) Systems Case Study. Front. Sustain. 2020; 1: 1–9.

41. Allaoui H, Guo Y, Sarkis J. Decision Support for Collaboration Planning in Sustainable Supply Chains. J. Clean. Prod. 2019; 229: 761–774.

42. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6: e1000097.

43. Caputo A, Kargina M. A User-Friendly Method to Merge Scopus and Web of Science Data during Bibliometric Analysis. J. Mark. Anal. 2022; 10: 82–88.

44. Aria M, Cuccurullo C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. J. Informetr. 2017; 11: 959–975.

45. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field. J. Informetr. 2011; 5: 146–166.

46. Sittón-Candanedo I, Alonso RS, Corchado JM, Rodríguez-González S, Casado-Vara R. A Review of Edge Computing Ref-erence Architectures and a New Global Edge Proposal. Futur. Gener. Comput. Syst. 2019; 99: 278–294.

47. Pullo S, Pareschi R, Piantadosi V, Salzano F, Carlini R. Integrating IOTA's Tangle with the Internet of Things for Sustainable Agriculture: A Proof-of-Concept Study on Rice Cultivation. Informatics. 2024; 11.

48. Krisnawijaya NNK, Tekinerdogan B, Catal C, van der Tol R. Reference Architecture Design for Developing Data Management Systems in Smart Farming. Ecol. Inform. 2024; 81: 102613.

49. Mohd Nizar NM, Jahanshiri E, Tharmandram AS, Salama A, Mohd Sinin SS, Abdullah NJ, Zolkepli H, Wimalasiri EM, Mohd Suhairi TAST, Hussin H, et al. Underutilised Crops Database for Supporting Agricultural Diversification. Comput. Electron. Agric. 2021; 180: 105920.

50. Talari G, Cummins E, McNamara C, O'Brien J. State of the Art Review of Big Data and Web-Based Decision Support Systems (DSS) for Food Safety Risk Assessment with Respect to Climate Change. Trends Food Sci. Technol. 2022; 126: 192–204.

51. Palocci C, Presser K, Kabza A, Pucci E, Zoani C. A Search Engine Concept to Improve Food Traceability and Transparency: Preliminary Results. Foods. 2022; 11: 1–13.

52. Kassahun A, Tekinerdogan B. BITA*: Business-IT Alignment Framework of Multiple Collaborating Organisations. Inf. Softw. Technol. 2020; 127: 106345.

53. Machado IA, Costa C, Santos MY Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. Procedia Comput. Sci. 2021; 196: 263–271.

FINANCING
This research was funded by MINCIENCIAS [Grant Number 934-2023], and the article processing charge (APC) was also supported by MINCIENCIAS [Grant Number 934-2023].

CONFLICT OF INTEREST
The author, William Alejandro Orjuela-Garzon, was employed by the company Inntegra SAS. The remaining authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: William Alejandro Orjuela-Garzón, Henry Cárdenas-Roa.
Methodology: William Alejandro Orjuela-Garzón.
Software: William Alejandro Orjuela-Garzón, Daniel Bustos-Vanegas, Juan Manuel Andrade-Navia.
Validation: William Alejandro Orjuela-Garzón, Henry Cárdenas-Roa.
Formal analysis: William Alejandro Orjuela-Garzón, Daniel Bustos-Vanegas, Juan Manuel Andrade-Navia.
Investigation: William Alejandro Orjuela-Garzón.
Resources: William Alejandro Orjuela-Garzón.
Data curation: William Alejandro Orjuela-Garzón.
Writing-original draft preparation: William Alejandro Orjuela-Garzón, Daniel Bustos-Vanegas, Juan Manuel Andrade-Navia.
Writing-review and editing: William Alejandro Orjuela-Garzón.
Visualization: William Alejandro Orjuela-Garzón.
Supervision: William Alejandro Orjuela-Garzón.
Project administration: William Alejandro Orjuela-Garzón.
Funding acquisition: William Alejandro Orjuela-Garzón.