AgriChain: Blockchain Syntactic and Semantic Validation for Reducing Information Asymmetry In Agri-Food

Pierluigi Gallo[1,2,3,*], Federico Daidone[2], Filippo Sgroi[4] and Mirko Avantaggiato[5]

[1]*Department of Engineering, University of Palermo, 90128 Palermo, Italy*

[2]*SEEDS s.r.l., academic spin-off of the Dept of Engineering at the University of Palermo, 90141 Palermo, Italy*

[3]*CNIT, Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Italy*

[4]*Department of Agricultural and Forestry Sciences, University of Palermo, Palermo, 90128, Italy*

[5]*former employee of SEEDS s.r.l., 90141 Palermo, Italy*

Abstract

Information asymmetry affects the actors of all the segments of the agri-food supply chain and can arise many problems in the market along the production chain. Transactions of agri-food products are asymmetric because suppliers and buyers have different levels of knowledge on the provenance, value, quality, and freshness of food. Collusive relations among the agri-food chain actors, especially between controllers companies and controlled ones, can cause market failures as they influence customers' purchase decisions and severe health accidents when food safety is compromised. This paper proposes using blockchain technology to combat information asymmetry and collusive relations. In addition to transparency, cryptography and trusts, which are natively provided by the blockchain, our approach provides a twofold mechanism for validating crowd sensed data: first, a lightweight syntax validation is run before writing data in the blockchain (providing accountability also thanks to immutability); then, a dedicated smart contract runs semantic validation in scenarios with multiple data sources. This semantic validation may reveal collusive behaviours, downgrade colluding nodes and exclude or down-weight their data in future validations. The smart contract seals data that pass both validations adding metadata on data quality. Results prove the feasibility of our solution on Hyperledger Fabric under the assumption that the majority of nodes are honest. Experimental results demonstrate that our implementation of the twofold validation using smart contracts scales well with the dimension of the blockchain state. Our mechanism may greatly impact Product Certification and Designation of Origin as it may be applied to check specific requirements for raw materials, products, and production processes and protect from the collusion of controlling consortia and certification bodies.

Keywords

Agri-food, economy, blockchain, smart contract, information asymmetry, validation

1. Introduction

In the globalised society, and above all in the developed economies, the quality and safety of agri-food productions have received increasing attention from the consumer as a result of the evolution intervened in recent years, in terms of production and marketing of products of

vegetable and animal origin in a fresh and transformed state. This structural and functional evolution of the sector is mainly due to some aspects. In fact, in the agri-food sector, there are new products with differentiated and differentiable characteristics, including agricultural commodities, vegetable and animal productions, highly-processed and high-service products.

Moreover, in the agri-food system, there is a strong integration of the productive sector with the final consumer market in terms of information flows, knowledge of markets, and consumer needs and expectations. The economic and technical literature reports the growing importance of food quality and safety. These concepts are related to brands, information transparency, traceability of production and commercial chains, the fight against counterfeiting, and food fraud [24, 9]. However, in countries with high per capita income, the current health and nutritional needs, expressed by new lifestyles, determine a rethinking of production protocols that are increasingly attentive to the problems of resource sustainability and the protection of environmental ecosystems and biodiversity. Finally, the continuous evolution of consumers' tastes and preferences expressed over time by the variations in the demand should not be overlooked. To manage this new scenario, the public operator and institutional figures have provided regulations at national and international levels, disciplinary and production controls, certifications and quality protection, international agreements, and trading platforms. However, modern technologies require the adoption of systems that can support themselves by minimising human intervention in data collection and certification processes. In this context, information availability becomes fundamental for consumers because they quest for valuable information to perceive and evaluate the quality of products, recognise the added value, and increase willingness to buy or pay more. For all these reasons, this paper aims to analyse the role of information and novel ICT technologies in creating higher standards of quality and improving the functional efficiency of agri-food production markets by reducing information asymmetries on the demand side.

The contribution of this paper joins together economy and computer science; first, we explain the economic and technical implications of the information asymmetry in the agri-food market, then explore a possible solution to reduce such an asymmetry using blockchain technology. Blockchain has intrinsic traits such as transparency, trust, and traceability; these features help to solve the information asymmetry but, alone, they are not enough to guarantee the data accuracy and validity. Blockchain technology provides data immutability, accountability, and traceability, but it does not guarantee the data quality. In the agri-food sector, data quality is the cornerstone; therefore, a blockchain-based AgriChain platform for data quality is necessary. Using blockchain and smart contracts and applying a novel data validation methodology, we combat information asymmetry and its negative influence on the net value of investments, the ranking of agri-food companies and their capability to access credit for financing their activities. AgriChain uses multiple data sources, in which data are analysed by a set of smart contracts implementing a two-step validation logic (syntactic and semantic). The **syntax validation works before data are written on the blockchain**; it checks both the data and the user's identity and guarantees the accountability of the written information. Then, AgriChain smart contract applies a **semantic validation that works after data are written on the blockchain and 'seals' them.** This validation smart contract fights information asymmetry, providing transparency and data accuracy.

The distortions of information asymmetry in the food market are described in Section 3. The

actors and the roles in the agri-food supply chain are discussed in Section 4. Section 6 describes the AgriChain methodology for validating and assessing data quality. Experimental setup and results are presented in Section 7, then related works and conclusions are drawn, respectively in Sections 8 and 9.

2. Background

This section briefly introduces the key elements of the proposed architecture, namely ontology and blockchain.

2.1. Ontology

In computer science, ontology is a way to represent semantics (the meaning) through the definition of categories, properties and relationships expressed through description logic [12]. An ontological approach enables or simplifies deductive reasoning, classification, problem-solving, and the simplification of information exchange among systems. Deductive reasoning is entrusted to the semantic reasoner, software capable of carrying out reasoning on formalised knowledge bases. It is capable of elaborating the knowledge base according to some rules to validate and analyse the knowledge base itself and, therefore, infer logical consequences. In 1999, the W3C adopted the Resource Description Framework (RDF), which became standard in 2004. RDF is a data model used to represent ontologies; the atomic data entity is the semantic triple, a set of three entities: subject-predicate-object. Triples represent a statement on semantic data (e.g., "Alice is 30", "Alice knows Bob"). SPARQL Protocol and RDF Query Language (SPARQL) is a SQL-like query language for receiving and manipulating RDF data. An implementation of SPARQL is included in Apache Jena, a Java framework for developing semantic web-oriented applications that include a SPARQL endpoint and supports a specific serialisation format named Turtle (Terse RDF Triple Language). RDF data validation is entrusted to Shapes Constraint Language (SHACL), which includes a list of constraints such as cardinality, range of values, etc. [7].

2.2. Blockchain

Blockchain is a distributed technology that allows for addition-only data storage. Each member of the distributed network (node) has its data replica on which it tracks every resource exchange (transaction) between participants. The transactions are grouped into blocks, linked together through a content hash, to form a chain. Members participate in the validation of transactions in order to add them to the blocks through a distributed consensus algorithm. There are several types of protocols, the most famous being Proof of Work (PoW), Proof of Stake (PoS), and Byzantine fault tolerance (BFT). Ethereum was the first blockchain platform that introduced smart contracts, small programs for validating transactions and performing the computation in a distributed way. Ethereum is a permissionless blockchain where anyone can participate in the network and participate in the consensus protocol.

Conversely, there are permissioned blockchains, such as Hyperledger Fabric (HLF), where participants need special permissions to be part of it. HLF is part of the broader Hyperledger

framework, which includes other distributed ledgers, libraries and tools, and the Linux Foundation supports it. Here the smart contracts are called chaincodes and enable to read (*query operation*) and write (*invoke operation*) the ledger. The ledger is included in a channel; nodes that participate in this channel can read, write and invoke smart contracts. An HLF instance can manage multiple channels and, therefore, multiple ledgers, defining different levels of scope for each node.

Since version 2.0, HLF supports chaincodes as an external service. In this case, the chain code management is independent of the node and allows us to define an endpoint where it is executed.[1] In this endpoint, we can also run more complex services, which the chaincode is capable of invoking, such as in [18] where external chaincodes are used to query external data sources. The call can be made in the single execution of the chaincode, or in case of longer processing times, the chaincode can exploit the oracle paradigm [5]. In this case, the chaincode emits an event that the service intercepts to start the computation of the request. When the service has finished the processing, it returns the output to the chaincode.

3. Information asymmetry and market distortion

From an economic point of view, it is well known the possibility to score the perceived quality of food products using a scale that spans from optimal to poor without interfering with its potential edibility. However, the hygienic and sanitary safety of the products to the final consumer markets is challenging to evaluate. Consumers have shown great interest in features defining food quality, thanks to an excellent spending capability and a more sensitive contest than in the past. Food quality is a multidimensional and dynamic concept [14]. Quality is "a complex value whose definition involves objective and subjective components. For this reason, quality is not a characteristic that can be immediately described or identified. However, it is an idea that each of us has concerning what we need to satisfy a specific need. The more the characteristics of a product correspond to the complex expectations we have concerning it, the more we will be inclined to consider its quality" [25]. It becomes essential to deepen the analysis on the perception of qualitative aspects, combining technical quality indicators with measures and models of customer satisfaction interpretation in the information economy's theoretical context. Indeed, placing on the market certified quality products is reflected in an increase in production costs and therefore in prices. Certification requires an estimation of the economic value attributed to the quality perceived by the customers and the evaluation of the premium price concerning the different and greater willingness to pay.

Information is an element that affects the functioning mechanisms of the markets, providing a twofold perspective. On the one hand, the "control" and the "management" of the information asymmetry between supply and demand, through the policy of trademarks, certifications, and labelling of agri-food productions. On the other hand, national and international public and private organisations and institutions preside over voluntary standardisation and establish rules and procedures for controlling market transaction costs. They check company and collective brands as precise quality signals, signals of value and contribute to strengthening

---

[1]Available at https://hyperledger-fabric.readthedocs.io/en/release-2.4/cc_service.html

the necessary operating conditions for the exchange, contributing to the reduction of the information asymmetry typical of imperfect markets [1, 20].

The quality of food production and the economic efficiency of the markets are closely connected and correlated to the growing role of information. This type of situation does not always safeguard the security and correctness of the information and the ability to choose given to informed consumers. From the point of view of the economic production efficiency of the markets, these elements contribute to creating a sort of functional distortions of the agri-food markets that can prevent their correct functioning under the profile of economic theory. These specific conditions seem to simultaneously produce disadvantages for producers and consumers in terms of the natural relationship between supply and demand, oriented to the balance of short and long term markets.

4. AgriChain actors and roles

The agri-food supply chain is composed of segments that cooperate to evolve the production process from field to fork. Information asymmetry typically manifests in the last segment of the supply chain affecting final customers but, in many cases, also influences other actors. The various segments concur to a holistic view of the good, including production and transformation processes. In case of partial or inaccurate information, two consecutive parts of the supply chain (e.g., production, transportation, transformation, stock) may experience information asymmetry too. For example, farmers know the history of the grain they grow - origin, timing, and treatments. This information may be hidden to the miller, whose knowledge is limited to storage in silos and the milling process. The same issue related to lack of knowledge occurs between miller and distributors and, more in general, in all the steps between different actors. The chain of value and responsibility that links those actors from farm to fork is affected by information asymmetry in all its links.

Farmers and industries need prompt and trusted information to make better decisions for growing or transforming agri-food products. The introduction of blockchain in the agri-food sector has represented a digital innovation aimed at increasing business income by reducing production inputs (and therefore of costs expressed at constant prices) and increasing the outputs (the quantity produced and therefore of revenues expressed at constant prices). Digital innovation is always aimed at increasing the company's competitiveness and technical and economic efficiency by optimising production factors and reducing variable costs. For example, accurate information on the state of plants brings to savings of water for irrigation, avoiding unnecessary wastes. The same happens for fertilisers and pesticides with knowledge on seasonal trends and infections. These decisions change the structure of production costs and positively affect the entrepreneur's net income.

The information asymmetry negatively influences production and marketing choices, and the potential problems along the supply chain may lead to market failure. An important issue is related to product certification about the designation of origin. Such certifications are characterised by strict requirements and are guaranteed by consortia and certification bodies. However, between the controlling and controlled entities may arise collusive relations, which are then difficult to discover and strongly affect the market. A recent example is given
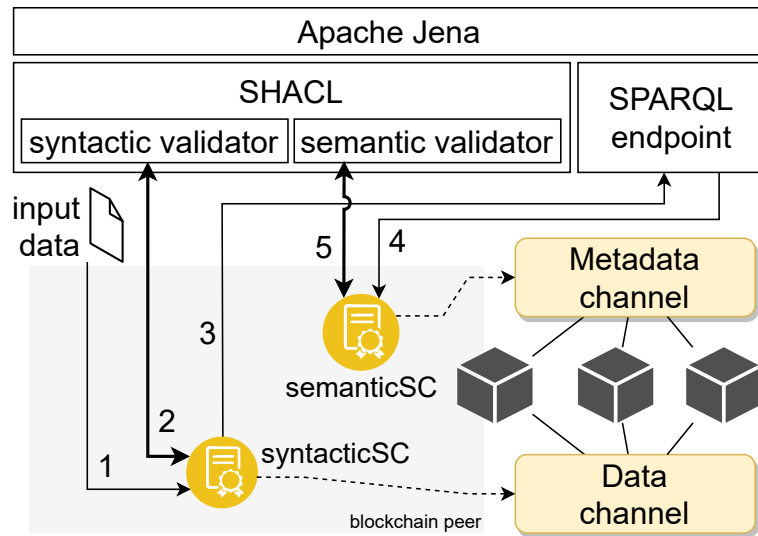
Figure 1: Blockchain node representation of our validation system. It includes a *SPARQL endpoint*, a *syntactic validator*, and a *semantic validator*

the production of ham under the Protected Designations of Origin (PDOs) "San Daniele" and "Parma", which require the use of a specific breed of pigs. However, a collusive system within the protection consortium eluded controls on the seed of the pigs and, in contrast with the production disciplinary, put on the shelves products whose PDO was not valid. The effects of information asymmetry apply both to product quality and health, as in the cases such as pistachio, whose origin has implications in terms of aflatoxin and ochratoxin and may cause risks for consumers' health [21]. This example shows that information asymmetry may have different facets. The consumer needs to know a product's provenance, but this information is not sufficient if it is not linked to the risks of products from a specific area.

5. Validation Architecture

To combat the information asymmetry, we provide AgriChain, a blockchain-based platform for semantic and syntactic validation that executes external smart contracts on HLF (see, Section 2.2). In this way, within a blockchain node, we can run complex services such as *Apache Jena*, a free and open-source Java framework for building semantic applications [3], otherwise impossible to be implemented as legacy smart contracts. It includes a *SPARQL endpoint*, i.e. *Fuseki*, and a *syntactic and semantic validator*, i.e. *SHACL*. As shown in Figure 1, each node runs the two smart contracts in yellow that interface with the services mentioned above. We use HLF channels to separate the essential information from metadata and facilitate their operations. Semantic validation works on datasets rather than on a single transaction. The traceability information is stored in the *data channel*, while the *metadata channel* is used to store useful elements for the validation operations.

5.1. Syntactic validation

The syntactic validation takes place before storing the information on the blockchain, implementing filtering on the single input data. The smart contract *syntacticSC* (see, 1 in Figure 1) receives as input the data and performs a signature validation. This step is shown in Algorithm 1, where *syntacticSC* takes *dataIn* as input parameter and passes it to *signatureValidation*() function (see, Line 2). If that check is successful, we continue with syntactic validation, calling *syntaticValidation*() function (see, Line 3), as described in Section 6, to invoke the *syntactic validator* service (see, 2 in Figure 1). This validation has to be customised as needed and depends on the context of the application, for example, to verify that a "*weight*" field has a numeric value expressed in *kg*. When the validation is successful, we map *dataIn* into a *dataOut* format (see, Line 4) valid to be loaded on *SPARQL endpoint* (see, 3 in Figure 1). We assume that the reference ontology is preliminary written on the blockchain and imported into the *SPARQL endpoint* before starting the data collection process. Writing the ontology on the blockchain guarantees interoperability and transparency in the definitions of products and links between them. Finally, we also store *dataOut* on data channel (see, Lines 5 and 6).

Algorithm 1 syntacticSC

Require: *dataIn* as input data
 1: **procedure** SYNTACTICSC(*dataIn*)
 2:     **if** SIGNATUREVALIDATION(*dataIn*) **then**
 3:         **if** SYNTACTICVALIDATION(*dataIn*) **then**
 4:             *dataOut* ← MAPPING(*dataIn*)
 5:             PUTSPARQL(*dataOut*)
 6:             PUTBC(*dataOut*)
 7:         **end if**
 8:     **end if**
 9: **end procedure**

5.1.1. Semantic validation

The semantic data validation process uses SHACL shapes, deriving from the ontology.[2] We assume that they are already present on the blockchain and used by the smart contract *semanticSC*. The *semanticSC*, as shown in Algorithm 2, receives as input the parameters *query*, that is, the SPARQL query which determines the subject of the validation, and *idShacl*, the identifier of a shape stored on the blockchain used in the validation. When this smart contract is invoked, it retrieves the *dataset* from the *SPARQL endpoint* (see, 4 in Figure 1), using *getSPARQL*() function with *query* parameter (see, Line 2). Similarly, we retrieves the *shacl* shape from blockchain with *getBC*() function (see, Line 3). Then we forward *dataset* and *shacl* shape to *SHACL validator* (see, 5 in Figure 1). Here, the *semanticValidator*() function calls the *semantic validator* service (see, Line 4) which performs a semantic validation and gives back the *result*. Now, *result*, along with

---

[2]We generated the SHACL shapes using the Astrea tool, https://astrea.linkeddata.es/. The shapes have been tuned, and we added missing validation elements from the ontology, such as the cardinality range.

dataset, are examined by a *calculateScore()* function (see, Line 5) which scores the validation performed. The single application defines the calculation of the score and its metric; for example, the closer the harvesting coordinates of different olives are, the more accurate the result that the crop belongs to an exact agricultural land. At the end, *query*, *idShacl*, *result*, and *score* are written on metadata channel as proof via *putBC*() function (see, Line 6).

Algorithm 2 semanticSC

Require: *query* as input query for SHACL validation
Require: *idShacl* as input id of SHACL shape
  1: **procedure** SEMANTICSC(*query, idShacl*)
  2:     *dataset* ← GETSPARQL(*query*)
  3:     *shacl* ← GETBC(*idShacl*)
  4:     *result* ← SEMANTICVALIDATOR(*dataset,shacl*)
  5:     *score* ← CALCULATESCORE(*dataset,result*)
  6:     PUTBC(*query, idShacl, result, score*)
  7: **end procedure**

The *reasonerSC* interfaces the blockchain with the reasoning service. When this smart contract is invoked, it queries the *SPARQL endpoint* (see 3 in Figure 1) to obtain the dataset to forward to the reasoner, indicated with 4). When the reasoner finishes its processing (see 5), the smart contract stores the result on the blockchain. If the result leads to new inferred triples from the initial dataset, the new data is updated in the *SPARQL endpoint* invoking *syntacticSC*. In such a case, the initial data are stored on the data channel, and the inferred information goes on the metadata channel.

6. AgriChain validation methodology

The agri-food sector includes multiple supply chains for the different agricultural products: tomatoes, wine, dairy, olive oil, etc. These supply chains involve many actors with different roles, and in most cases, they hold contrasting interests. Agricultural entrepreneurs, transformation industries, transport, logistics, and great and small distribution are exemplary actors that appear in many agri-food supply chains. However, any chain has its peculiar actors with specific needs and roles. For example, in the simplified model of the olive oil supply chain shown in Figure 2, there are farmers, olive growers' cooperatives, warehouses, shops, and customers as the main actors. These actors typically provide data through human operators, which are not trusted by default. To solve the problem of mistrusted operators, the authors propose to use IoT devices. However, this strategy shifts the point of trust from humans to IoT devices. IoT sensors are owned and maintained by those actors indicated above and can be maliciously manipulated according to their specific interests. To guarantee data quality, AgriChain leverages the double validation indicated above, invoking dedicated smart contracts.

The input syntactic validation, performed by smart contract *syntacticSC* (see, Pseudocode 1), checking that the transaction contains specific fields, as exemplary shown in Listing 1, including the actor's signature. The smart contract checks multiple signatures if multiple actors are
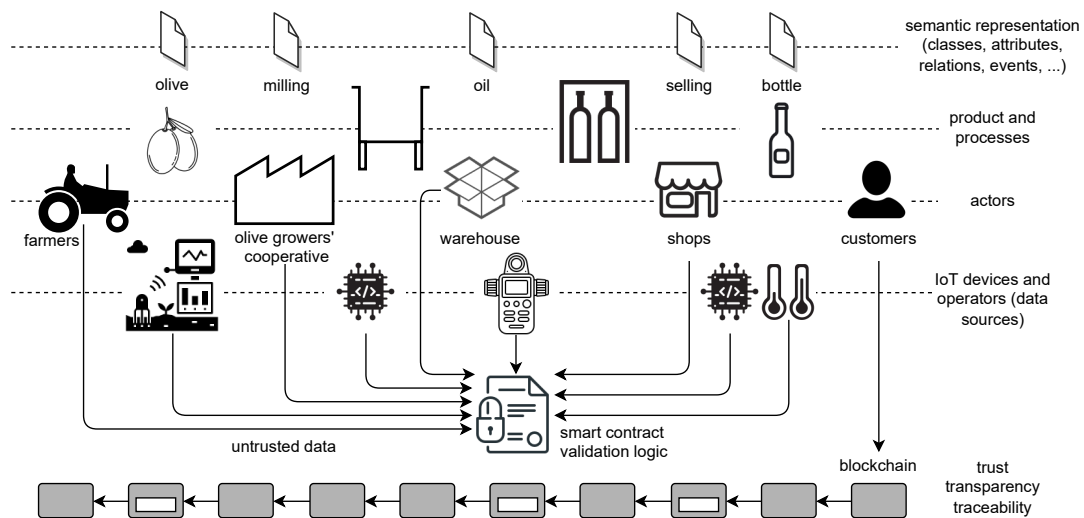
Figure 2: Data sources for AgriChain and smart contracts for data validation. Syntax-validated transactions are in gray, those semantically validated are in white.

involved in the transaction. This validation is performed on a transaction *before* being written on the blockchain. This preliminary validation guarantees accountability because each piece of data is linked to an accountable entity, but still, it does not protect from the 'garbage in, garbage out' problem. In other words, this lightweight syntax validation checks the identity of the data provider, the timestamp, and other metadata without guaranteeing 'semantic' validity.

```
{
    "actor": {
        "signature": "ebf3d6a0e54d249ff..." },
    "res_details:": {
        "res_name": "olives01@field01",
        "hasGeoTag": true,
        "hasWeight": true },
    "data": {
        "lat": 38.120240,
        "lon": 13.357388,
        "kg": 10 },
    "ts": "2020-05-30T16:06:44+01:00"
}
```

Listing 1: Syntactic validation - Fields extracted from the transaction.

The second check involves both syntactic and semantic validation; in what follows, we stress the semantics aspects. Here, the smart contract *semanticSC* (see, Pseudocode 2) takes care of the validation on a more extensive set of data that, grouped, have a special meaning; the validation logic depends on the specific supply chain and the meaning of data, in our experiments we focused on the geographical origin of the olive oil product. Unlike the typical blockchain validation, our semantic validation is performed *after* the data is written on the blockchain, it is
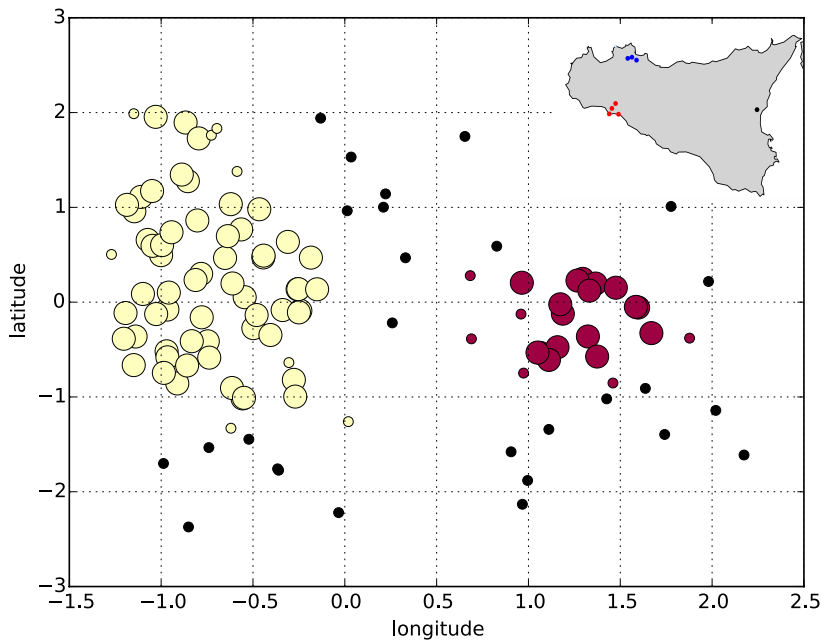
Figure 3: Exemplary semantic validation for geo coordinates of extra-virgin olive oil origin provided by 120 actors. The majority cluster (67 dots) is in yellow, noisy measures in black (30 elements), colluding nodes (23) in red.

triggered by new data arrivals that are semantically linked to the previous ones. For example, the geographic coordinates provided by several harvesting operators through their smartphones and IoT devices with GPS receivers are in Listing 1, providing the location of the product and farm *field01*. As shown in Figure 3, the syntax validation smart contract uses clustering to estimate the position (the mean of the majority cluster) from malicious and colluding nodes (in red).

6.1. Costs and benefits of the proposed solution

When an agri-food related business choices to use blockchain technology to implement its food supply chain in some or all aspects, it is choosing to undergo some change. Change is not always good for business, so why should a business decide to switch to a blockchain-based solution? Because using blockchain expresses the company care about transparency, thus inspiring old customers to possibly buy more products and/or new ones to switch from another brand to this one. Of course, every kind of IT infrastructure comes with costs of installation and maintainability. We propose those costs to be proportionally assigned to the *n* involved actors. This solution could be thought of as a blockchain-based pay-per-use like a subscription system.
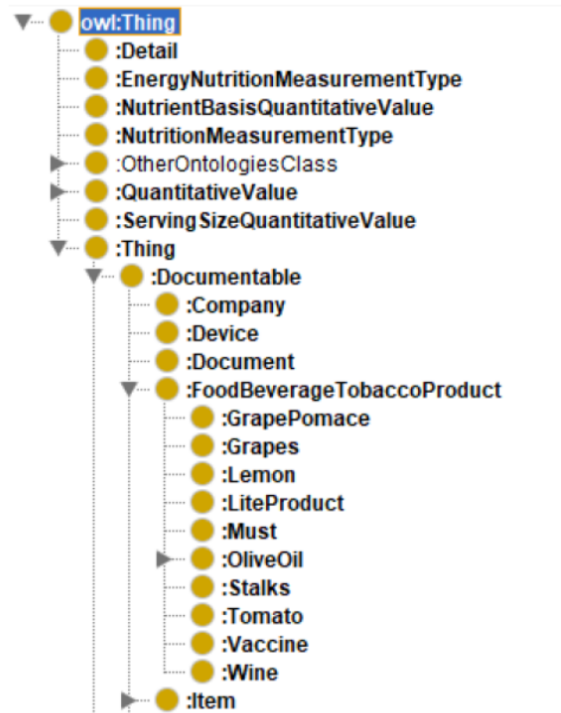
Figure 4: Protégé class hierarchy overview for *sb:OliveOil*.

7. Experimental Setup and Results

Part of the platform presented in this paper was proposed within the *DEMETER*[3] project, which leads the digital transformation of the European agri-food sector through the rapid adoption of advanced IoT technologies, data science and smart farming, ensuring its long-term viability and sustainability. Our blockchain currently runs within the *DEMETER* ecosystem, and the project partners can invoke its services. A fundamental part is the semantic model, used as a common language between different project entities. It is based on the GS1 vocabulary, extended, revised and refined to be able to describe an entire supply chain. We exemplary show the olive oil supply chain (see, Figure 4), where we have extended the *gs1:FoodBeverageTobaccoProduct*[4] class with *sb:OliveOil*[5] to be able to map the entire process. In addition to the interoperability offered by the semantic model and its mappings with other ontologies, the platform offers APIs compliant with the OpenAPI standard. Seeing the generality of the platform, we, as a case study, have implemented the validation of olive harvesting in the olive oil supply chain. Within the *SHACL validator*, we have added a clustering algorithm, the *DBSCAN* [19], to calculate the

---

[3]Available at https://h2020-demeter.eu/
[4]Available at: https://www.gs1.org/voc/FoodBeverageTobaccoProduct
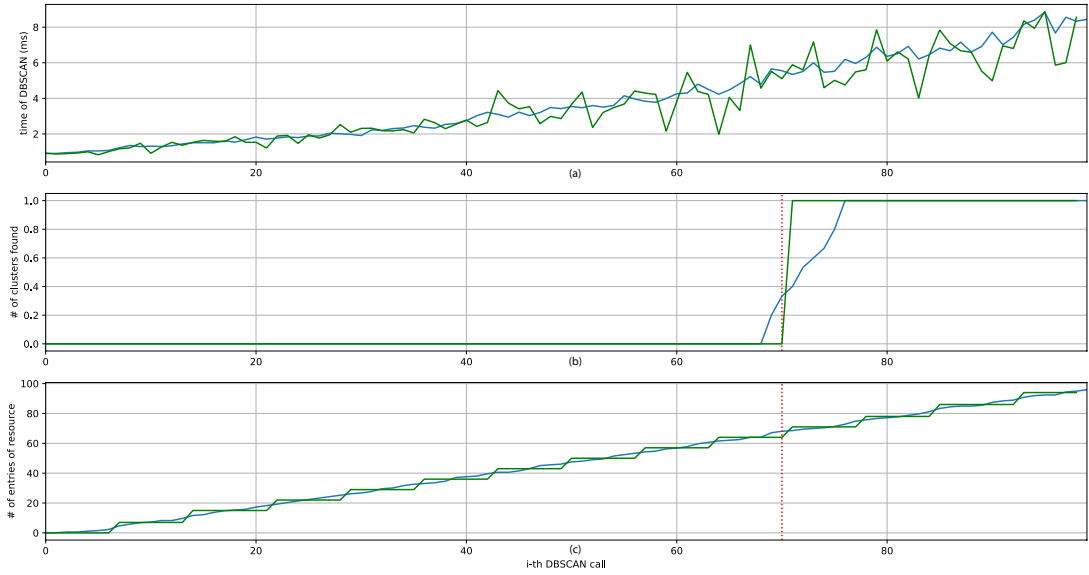[5]Available at: https://seedsbit.com/ontology/#OliveOil

Figure 5: Time spent by the chaincode for clustering geographical points with DBSCAN for 100 consecutive invokes (a); Average number of points to create at least one cluster (in red) (b); Number of entries written on the blockchain (c). One typical run is depicted in green; the average value on 15 runs appears in blue. When a new pair of coordinates is added, the smart contract is triggered; the ith call works on a bigger state than the $i − 1$th.

proximity of the harvested olives to the soil. Our blockchain platform of choice to illustrate our work is Hyperledger Fabric, although the SeedsBit platform uses multiple blockchain platforms, including MultiChain and Ethereum.

As introduced in Section 6, we used Hyperledger Fabric to implement our model partly and to give some experimental results in terms of performances. Our test network was composed of two Fabric organisations, having two peers each. Moreover, we used the RAFT algorithm [17], which is the default consensus protocol for Hyperledger Fabric. RAFT is a CFT (Crash-Fault Tolerant), but it can be easily substituted with a BFT (Byzantine Fault Tolerant) as Fabric has a modular approach to the consensus protocol [4].

Thus we had five nodes running for consensus purposes. The blockchain was deployed on a single host configuration on a machine with the following specs: Intel© Xeon© CPU E5-1660 v3 @ 3.00GHz with 32 gigabytes of RAM. Figure 5 shows, out of 100 consecutive invocations of the smart contract *semanticSC*, the time spent by the *DBSCAN* algorithm for clustering (see, Figure 5a), the number of clusters found (see, Figure 5b), and the number of entries used by *DBSCAN* (see, Figure 5c). At each invocation, we assume that the number of entries has increased by 1 unit, so *syntacticSC* has inserted a new entry into the blockchain. We can see how the analysis of 100 points, the most computationally expensive part, uses about 8 ms, with is compatible with the smart contract execution. The clusterisation of the terrain, with about 75 points, required 6 ms.

114

8. Related work

The problem of information asymmetry in food traceability has multiple facets that have been traditionally tackled singularly and using old paper documents and product specifications. Our approach towards information asymmetry is to improve transparency under multiple points of view: economy, blockchain technology, data quality.

From an economic point of view, it is well known the possibility to score the perceived quality of food products using a scale that spans from optimal to poor without interfering with its potential edibility. However, the hygienic and sanitary safety of the products to the final consumer markets are challenging to evaluate. Consumers have shown great interest in features defining food quality, thanks to a greater spending capability and a more sensitive contest than in the past. Food quality is a multidimensional and dynamic concept [14]. Quality is a complex feature made by objective and subjective components. For this reason, quality cannot be immediately described or identified, but it is a subjective idea that involves personal needs. The more the characteristics of a product match our expectations, the more we will be inclined to consider its quality [25]. It becomes important to deepen the analysis on the perception of qualitative aspects, combining technical quality indicators with measures and models of customer satisfaction interpretation in the information economy's theoretical context. Indeed, placing on the market certified quality products is reflected in an increase in production costs and therefore in prices. Certification requires an estimation of the economic value attributed to the quality perceived by the customers. This requires the evaluation of the premium price concerning the difference and greater willingness to pay. Information is an element that affects the functioning mechanisms of the markets, providing a twofold perspective. On the one hand, the "control" and the "management" of the information asymmetry between supply and demand, through the policy of trademarks, certifications and labelling of agri-food productions. On the other hand, national and international public and private organisations and institutions preside over voluntary standardisation and establishing rules and procedures for controlling market transaction costs. Company brands, collective brands, signals of quality and value work as media communication and contribute to strengthening the operating conditions necessary for the realisation of the economic exchange, contributing to the reduction of the information asymmetry typical of imperfect markets) [1, 2, 20].

From the point of view of the economic efficiency of the product markets, these elements contribute to creating a sort of functional distortions of the agri-food markets that can prevent their correct functioning under the profile of economic theory. These specific conditions seem to simultaneously produce disadvantages for producers and consumers in terms of the natural relationship between supply and demand, oriented to the balance of short and long term markets. In fact, in [22, 10, 23, 13] many different ways to leverage blockchain technology in this direction are illustrated. In [22] it is explained why a food traceability system based on RFID and blockchain would be ideal in China after many food safety accidents happened. These accidents were related to inadequate and primitive food supply chain management. In [10], the typical steps and places of a blockchain-based food traceability system are shown. The authors of [8] conclude their work stating that 'there are still few uses to support that some properties of blockchain implementation might be useful towards supply chain management'. In [13], it is reported how Walmart - one of the biggest American corporations in the hypermarket's

field - in collaborations with IBM, reduced the time needed to track the origins of mango "from seven days to 2.2 seconds". These performances also show how blockchain is, without doubt, a solution to at least consider when talking about food safety and food supply management. The blockchain used in this pilot study was Hyperledger Fabric. Among others, we found the high customisation possibilities offered by Hyperledger Fabric and its growing community and scientific literature response and usage. We see in [11] that performance is not going to be an issue at least in terms of transactions/second (the authors state that - after heavy re-engineering - they reached 20000 transactions/second). On the other hand, in [16] we see possible problems in critical scenarios if the blockchain physical network undergoes latency. In addition to performance, the blockchain has been used for guaranteeing high-quality data [26, 15].

9. Conclusion and Future Work

Quality of food production and the economic efficiency of the markets are closely connected and correlated to the growing role of information. This type of situation does not always safeguard the security and correctness of the information and the ability to choose given to informed consumers. The central role of the agri-food sector requires quality of data because erroneous, malicious, and missing information affect the food supply chain in terms of quality and safety. This paper presented AgriChain as a mechanism for validating data syntactically before being included in the blockchain and semantically before being sealed. These two validations are executed through a distributed logic, implemented with one or more dedicated smart contracts. Typically the blockchain is the preferred technology when seeking trust, transparency and traceability among actors who do not trust each other or have contrasting interests. We demonstrated how AgriChain goes beyond this vision on data management, breaking the simplistic concept that data written on the blockchain are trustful because they have been validated in advance. Indeed, Agrichain performs only a lightweight validation before including the information into a block; this only guarantees accountability and syntax consistency. From the semantic point of view, the second validation guarantees first data cleaning second data quality assessment. AgriChain performs *data cleaning* applies clustering algorithms implemented as smart contracts on data collected through crowd-sensing. Standard data cleaning methods aim at detecting and removing repeated entries, detecting outliers, checking data volumes. In general, such methods do not deal with malicious data sources. Then, AgriChain smart contract checks accuracy, timeliness, completeness, uniqueness, and consistency [6] and provides KQIs, *Key Quality Indicators* which are added, as metadata, as a data seal on the blockchain. This paper presented a new methodology for using smart contracts to enforce a twofold validation and guaranteeing the quality of data for food traceability.

Acknowledgments

References

[1] George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. In *Decision Science*, pages 261–273. Elsevier, 2017.

[2] Gervasio Antonelli. *Unione Europea, qualità agro-alimentare e commercio mondiale. Opportunità e minacce per i prodotti tipici delle Marche.* QuattroVenti, 2001.

[3] Apache. A free and open source java framework for building semantic web and linked data applications, 2022.

[4] Artem Barger, Yacov Manevich, Hagar Meir, and Yoav Tock. A byzantine fault-tolerant consensus library for hyperledger fabric, 2021.

[5] Abdeljalil Beniiche. A study of blockchain oracles, 2020.

[6] Hongju Cheng, Danyang Feng, Xiaobin Shi, and Chongcheng Chen. Data quality analysis and cleaning strategy for wireless sensor networks. *Eurasip Journal on Wireless Communications and Networking*, 2018(1), 2018.

[7] Ben De Meester, Pieter Heyvaert, Dörthe Arndt, Anastasia Dimou, and Ruben Verborgh. Rdf graph validation using rule-based reasoning. *Semantic Web*, 12(1):117–142, 2021.

[8] S. Matthew English and Ehsan Nezhadian. Application of Bitcoin Data-Structures & Design Principles to Supply Chain Management. *arXiv preprint arXiv:1703.04206*, 2017.

[9] Huanhuan Feng, Xiang Wang, Yanqing Duan, Jian Zhang, and Xiaoshuan Zhang. Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges. *Journal of cleaner production*, 260:121031, 2020.

[10] Juan F. Galvez, J. C. Mejuto, and J. Simal-Gandara. Future challenges on the use of blockchain for food traceability analysis. *TrAC - Trends in Analytical Chemistry*, 107:222–232, 2018.

[11] Christian Gorenflo, Stephen Lee, Lukasz Golab, and Srinivasan Keshav. FastFabric: Scaling Hyperledger Fabric to 20,000 Transactions per Second. In *ICBC 2019 - IEEE International Conference on Blockchain and Cryptocurrency*, pages 455–463, 2019.

[12] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In *Handbook on ontologies*, pages 1–17. Springer, 2009.

[13] Reshma Kamath. Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM. *The Journal of the British Blockchain Association*, 1(1):1–12, 2018.

[14] Kelvin J. Lancaster. A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2):132–157, 1966.

[15] Danwei Liang, Jian An, Jindong Cheng, He Yang, and Ruowei Gui. The quality control in crowdsensing based on twice consensuses of blockchain. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 630–635, 2018.

[16] Thanh Son Lam Nguyen, Guillaume Jourjon, Maria Potop-Butucaru, and Kim Loan Thai. Impact of network delays on Hyperledger Fabric. In *INFOCOM 2019 - IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2019*, pages 222–227, 2019.

[17] Diego Ongaro and John Ousterhout. In search of an understandable consensus algorithm. In *2014 USENIX Annual Technical Conference ATC 14)*, pages 305–319, 2014.

[18] Srinath Perera, Amer A Hijazi, Geeganage Thilini Weerasuriya, Samudaya Nanayakkara, and Muhandiramge Nimashi Navodana Rodrigo. Blockchain-based trusted property transactions in the built environment: Development of an incubation-ready prototype. *Buildings*, 11(11):560, 2021.

[19] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[20] Joseph E Stiglitz. The causes and consequences of the dependence of quality on price. *Journal of economic literature*, 25(1):1–48, 1987.

[21] Seyedeh Faezeh Taghizadeh, Ramin Rezaee, Gholamhossein Davarynejad, Javad Asili, Seyed Hossein Nemati, Marina Goumenou, Ioannis Tsakiris, Aristides M Tsatsakis, Kobra Shirani, and Gholamreza Karimi. Risk assessment of exposure to aflatoxin b1 and ochratoxin a through consumption of different pistachio (pistacia vera l.) cultivars collected from four geographical regions of iran. *Environmental toxicology and pharmacology*, 61:61–66, 2018.

[22] Feng Tian. An agri-food supply chain traceability system for China based on RFID & blockchain technology. In *2016 13th International Conference on Service Systems and Service Management, ICSSSM 2016*, pages 1–6. IEEE, 2016.

[23] Feng Tian. A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things. In *14th International Conference on Services Systems and Services Management, ICSSSM 2017 - Proceedings*, pages 1–6. IEEE, 2017.

[24] S. Vieri. Quality Products and Genetically Modified Organisms in Italy: Hazards and Possibile Enhancements. *Journal of Nutritional Ecology and Food Research*, 1(1):68–77, 2013.

[25] S. Vieri. Conflitti di maniera e accordi di sostanza, 2015.

[26] Jingzhong Wang, Mengru Li, Yunhua He, Hong Li, Ke Xiao, and Chao Wang. A Blockchain Based Privacy-Preserving Incentive Mechanism in Crowdsensing Applications. *IEEE Access*, 6:17545–17556, 2018.