# An Agri-Food Data Platform for Food Safety and Certification

Nikos Papageorgopoulos
*Robotics & Cognitive Systems Unit*
*UBITECH*
Athens, Greece
npapageorgopoulos@ubitech.eu

Danai Vergeti
*Robotics & Cognitive Systems Unit*
*UBITECH*
Athens, Greece
vergetid@ubitech.eu

Elena Politi
*Robotics & Cognitive Systems Unit*
*UBITECH*
Athens, Greece
epoliti@ubitech.eu

Dimitris Ntalaperas
*Robotics & Cognitive Systems Unit*
*UBITECH*
Athens, Greece
dntalaperas@ubitech.eu

Xanthi S. Papageorgiou
*Robotics & Cognitive Systems Unit*
*UBITECH*
Athens, Greece
xpapageorgiou@ubitech.eu

Manos Karvounis
*AGRROKNOW*
Athens, Greece
manos.karvounis@agroknow.com

Giannis Stoitsis
*AGRROKNOW*
Athens, Greece
stoitsis@agroknow.com

Braminir Rakic
*PROSPEH*
Belgrade, Serbia
branimir.rakic@origin-trail.com

Milos Kotlar
*PROSPEH*
Belgrade, Serbia
milos.kotlar@origin-trail.com

Simeon Petrov
*SIRMA AI*
Sofia, Bulgaria
simeon.petrov@ontotext.com

Antoniy Kunchev
*SIRMA AI*
Sofia, Bulgaria
antoniy.kunchev@ontotext.com

***Abstract*—**
Food safety is undergoing through tremendous challenges over the last years, with food scandals and contamination issues putting constant pressure to global markets, while consumers demands for high quality of products are increasing. This raises the need for increasing stakeholders' knowledge of the food production process and adopting data sharing practices in the product and supply chain management. Data sharing platforms can undertake the role of creating high value from data while facilitating secure and mutually beneficial multi-partner data sharing. Our proposed system aims to deliver an industrial data platform that will facilitate the exchange and connection of data between different food safety actors, who are interested in sharing information critical to certification, while boosting the way that food certification takes place in Europe.

***Index Terms*—food safety, data platform, certification**

## I. Introduction

Over the last years, we have witnessed major changes in the food sector, with a tremendous emphasis being put on food safety. A series of food safety scandals and health incidents, such as the Mad cow disease in the 90s [1], or the horse meat scandal of 2013 [2] have led into the international alignment of food safety standards through the Global Food Safety Initiative (GFSI) [3]. Governments also apply stricter policy and legislation, such as the integrated Food Safety policy of the European Commission2 [4] and the US Food Safety Modernization Act (FSMA3) [5]. Meanwhile, there exists an increased pressure for the agri-food and grocery sector to ensure that their suppliers comply with food safety standards that are recognised by the GFSI. This translates into more pressure for all stakeholders in the supply chain to exchange data critical to food safety assessment and assurance in a timely, trusted secure manner [6].

The gap between producers and consumers, as well as the lack of consumers' knowledge and control of food production in global modern food systems with long supply chains can be addressed by increasing consumers' knowledge of the food production process and adopting data sharing practices in the product and supply chain management [7]. This can be achieved through certification schemes, and food safety standards [8] or licenses that clearly state the data reuse conditions, thereby creating legal clarity for the researchers who reuse the data [9]. Despite of the inherent benefits of data sharing, there is evidence that data exchange and reuse practices may be limited. In this respect, mutual adoption between stakeholders is a contentious issue, with data producers (a.k.a., subjects) being concerned about how their data is being used or misused. Moreover, trust is an important factor that impedes

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00203

the sharing data among stakeholders of the supply chain [9]. In this context, data sharing platforms can undertake the role of creating high value from data that can facilitate secure and mutually beneficial multi-partner data sharing and encourage user participation in circumstances where user engagement is needed [7].

Our work aims to deliver an industrial data platform that will significantly boost the way that food certification takes place in Europe. It brings together and builds upon existing innovations from innovative ICT SMEs to deliver a uniquely open and collaborative virtual environment. The platform will facilitate the exchange and connection of data between different food safety actors, who are interested in sharing information critical to certification by delivering is going to be an open, distributed and innovative data-driven platform that aspires to catalyse the digital evolution of the quite traditional but very data-intensive business ecosystem of global food certification.

## II. Related Work

Data sharing has proven to be central and a valuable strategic resource for achieving competitive product delivery, elevation of digital platform business models and enhancement of operational efficiencies [10]. Data sharing platforms can create value for the participant stakeholders from collecting, integrating, and sharing different types of data. They can be classified in three general categories, a) Personal Data Platforms for the collection and management of personal information, such as the PIMCITY [11], where information of interest is selected, classified and assessed in terms of privacy and personal data management in order to increase transparency and provide citizens, organizations and companies control over their data, or the KRAKEN which facilitates the production/reparation and quality control of functional parts in the area of hybrid manufacturing [12], b) Industrial Data Platforms such as DataPorts [13] which connects existing digital infrastructures of seaports and their systems and sets rules on safe and reliable data sharing and trading with powerful services of data analytics, or the i3 Market [14], an intelligent, interoperable, integrative and deployable open source MARKETplace with trusted and secure software tools for incentivising the industry data economy, and c) Mixed Data Platforms such as the TRUSTS platform [15] that aims to reinstate trust previously placed in the data market.

In the food safety sector, there is a need to represent all food safety standards and their specifications for data monitoring and collection as commonly referenced and interoperable information models that can link, map, translate and transform different data formats in equivalent versions and formats [16]. Blockchain technology is considered as a promising technology that can help to build trust mechanisms for solving the transparency and security issues through the full information transparency and security dimension of food chains [16]. Blockchains, which are inherently distributed by design, are considered an innovation tool that is predicted to add the most value to agri-food supply chains [17].

In this work, we propose an open, shared, collaboratively developed and evolved platform that aims to open new directions for the management and operation of a marketplace via innovative services that combine, enrich and serve heterogeneous data sources, types and formats. This will bring competitive advantages to all food sector businesses that demand easy, fast, and actionable access to variegating food safety data from multiple devices and in various settings (on-site access and recommendations, responsiveness and adaptability in changes, etc.)

The contribution of the proposed system is as follows:

- An Attribute-Based Encryption technique that encrypts data using an access policy which specifies the attributes a user should be entitled to before being able to decrypt a file. A great advantage of this approach is that the owner can encrypt data based on desired attributes set, therefore allowing more fine-grained control.
- Authorization technique when accessing resources throughout the proposed platform.

## III. Agri-food Data Platform

### A. Architecture

The proposed architecture is a loosely-coupled modular architecture that provides enhanced flexibility in order to adapt and connect the various components that will be implemented as software modules, as depicted in Fig. 1. The major focus was on the functional decomposition, the strict separation of concerns, the dependencies identification and especially the data flow. As such, each component has been designed with the aim of delivering specific business services with a clear context, scope and set of features.

Our system provides a scalable and flexible environment with respect to interoperability of the various components that are facilitating the execution of analytics, data monetization and sharing through secure, transparent and advanced functionalities and features. To achieve this, all components of the our architecture provide well-defined interfaces to ensure the seamless integration and operation of the integrated platform.

The system architecture consists of a set of loosely coupled architectural components which are organized in three logical architectural layers:

- The data curation and enrichment layer which includes all the components which participate mainly in data ingestion, preparation, semantic enrichment and maintenance processes.
- The core services and backend data platform are the components which make use of the data stored and exchanged into the platform and perform the main data processing, encryption-decryption, analysis, identity and monetization services.
- The applications and marketplace layer includes the final offered services of system as they are implemented and provided by the lower architectural layers.

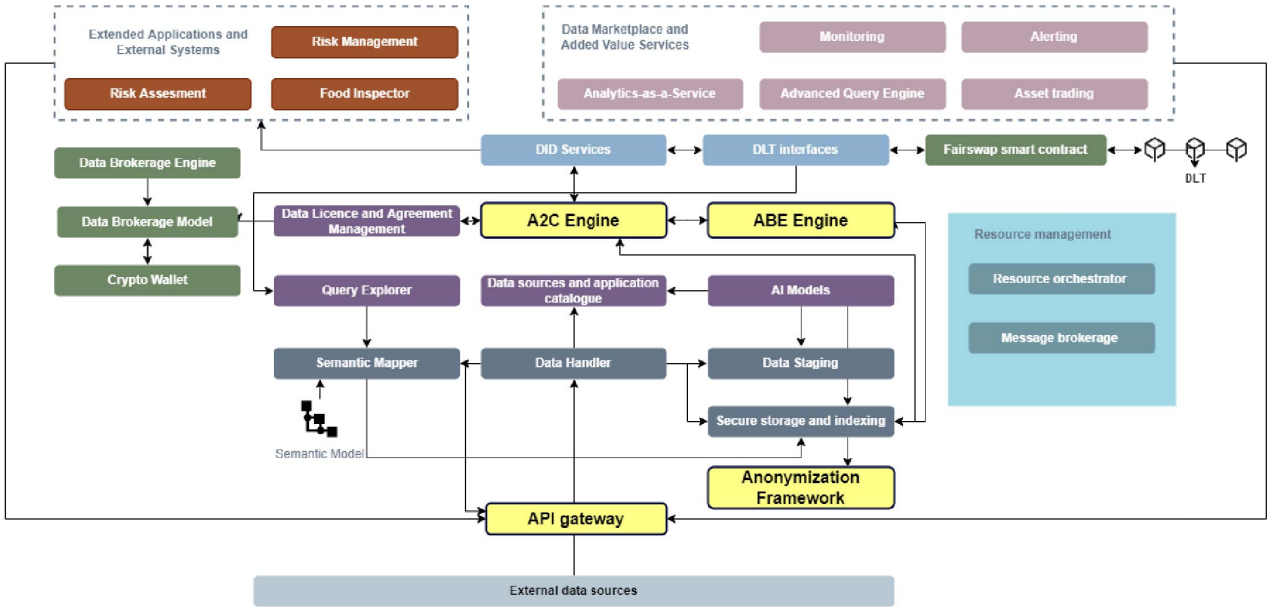


Fig. 1. Agri-Food Data Sharing Platform Architecture.

### B. Description of components

Supporting the everyday transactions as well as the data asset trading in food safety and certification requires the harmonization of multidisciplinary data deriving from a number of heterogenous data sources. In this section we provide a description of the components that comprise the system architecture. These facilitate the execution of analytics, data monetization and sharing through secure, transparent and advanced functionalities and features.

*1) Data Handler:* The Data Handler ingests the data and performs data ingestion functionalities for collecting and storing (aggregated) data from various data streams. The Data Handler performs Extract, Transform, Load (ETL) processes and implements a first lever of data transformation regarding a set of supported standards (WoT, GS1).

*2) Data Staging:* The Data Staging component consists of data management systems in regards to the stored data types. The collected data is provided on batches or collected and ingested by the Data Handler and are ready for semantic enrichment by the Semantic Mapper.

*3) Semantic Mapper:* The Semantic Mapper provides semantic enrichment of the data using the Semantic Model and generates the relevant Resource Description Framework (RDF) representation of the data which is stored in the Secure storage and Indexing.

*4) Secure storage and Indexing:* This component contains the semantic repository of the project where all necessary knowledge for running the platform is persisted. It is capable of storing and managing large amount of data in structured or unstructured format, as well as, semantic repositories (GraphDB) for the storage of the knowledge graphs generated and used in the platform.

*5) Query Explorer:* The component provides various interfaces for accessing the users' stored data in the platform, their semantic representation and linking to various ontologies and data sources in a consistent and unified manner. It interacts with services in data curation and semantic enrichment components helping the users to focus on the semantic instead of struggling with various data sources specifics. It also provides a way for interchanging data between platform users featuring data market functionality and allowing data consumers to use the platform in a data source independent manner.

*6) Data Sources and Applications catalogue:* The Data Sources and Applications catalogue is a set of extensions to the platform Semantic Object Model Language (SOML) schema [18]. Each data object which can be retrieved from an external source and more specifically the data type is declared as such an extension as well as the service which must be called to retrieve it. It consists of two parts:

a. Set of microservices (API calls) wrapping the external data sources and providing the Semantic Mapper (Apollo Federations Service) with all the needed data. If any data preprocessing is needed, it will be implemented in the corresponding microservice.
b. Set of definitions – extensions to the SOML representing the structure of data pieces retrievable from the corresponding (remote) data source and their mapping to semantic objects.

*7) Data Licence and Agreement Management:* The Data License and Agreement Manager is the component responsible for handling all processes related to the data licenses and

IPR attributes, as well as enabling the drafting, signing, and enforcing the smart data contracts that correspond to data sharing agreements between platform users. This component is provisioned to handle the data exchange and data transformation between the data curation and semantic enrichment layer (Data Staging), the automated contract negotiation and monetization layer and the Access and Authorization Control Engine.

*8) AI Models:* A number of AI-powered models and algorithms enhance the processing, forecasting and predictive capabilities of the platform, so that its users may generate more value from the data assets they use.

*9) Distributed Ledger Technology (DLT):* The data access and brokerage mechanisms are supported by state-of-the-art decentralized identity management provided by the DID Services. This component ensures provision and resolution of the Decentralized Identifier Descriptor (DID) and the relevant verifiable credentials of each organization that wants to perform any action on the data (provision, request, update) using Distributed Ledger Technology (DLT). The DLT Interfaces offers an abstraction and data management layer over DLT and facilitates the communication among the DID Services, the A2C Engine, and the Secure Storage and Indexing in order to manage traceability data exchanges through the platform, as well as, transparency and immutability of the data transactions.

A DLT is a replicated database that is consensually shared and synchronized with a specific protocol across multiple sites, institutions, or geographies, components of which are typically owned and accessible by multiple entities. The key property of DLT technology is that it is "decentralized", meaning being maintained in such a way that no central service or authority is needed to operate the system and broker transactions between participants.

One type of DLT technology is blockchain, named by the specifics of the protocol by which data is replicated and shared between DLT stakeholders (as a series of linked, cryptographically verifiable blocks of data, hence block-chain). DLTs key characteristics are achieving high resilience and increased data integrity, which is why it has been utilized in many enterprise use cases such as supply chain visibility and trade finance.

## IV. Security and access control Components

In the following subsections we describe the important technologies which are utilized in order to ensure proper authentication and authorization when accessing resources throughout the proposed platform. The first, Attribute-Based Encryption (ABE) addresses the issue of encrypting documents and data according to a set of attributes, while Attribute-Based Access Controller (ABAC) addresses the authorization aspect of accessing resources, based on both environmental and user-specific attributes. These core technologies are analyzed in the following subsection.

### *A. Attribute-Based Encryption Engine (ABE)*

The ABE in the proposed architecture is not used directly for encrypting records. The ABE is a promising new technique used to encrypt data without having to know the users beforehand [19]. The idea is that it encrypts a file using an access policy which specifies the attributes a user should be entitled to before being able to decrypt a file [20]. A great advantage of this approach is that the owner can encrypt data based on desired attributes set, therefore allowing more fine-grained control. Moreover, the ABE mechanism is scalable meaning that it offers symmetric key encryption, while it allows multiple independent authorities to issue attributes.

### *B. Attribute-Based Access Controller Engine (ABAC)*

The ABAC is an Authorization Access Control Engine that provides the access control mechanisms within our proposed platform [21]. ABAC does not require the use of any key system and it is a design concept which is used to control access to "objects" based on object attributes. In our implementation the access rights are granted to users through the use of policies in which attributes are combined together. The differentiation of the ABAC is the concept of policies in which multiple different attributes are evaluated through a complex Boolean rule set. As such, the model supports Boolean logic, in which rules contain "IF, THEN" statements about who is making the request (subject), the resource (object) and the action (operation).

The ABAC and ABE are policy-based, combined and they secure cloud persisted health data. The ABAC layer provides a fine-grained access control by evaluating rules, while the ABE layer authorizes access by decrypting the symmetric key.

### *C. Authentication*

Technologically, there are well-known industry standards enforcing this, such as OAuth and and JSON Web Tokens (JWTs).

OAuth is an open standard for access delegation, commonly used as a way for Internet users to grant websites or applications access to their information on other websites but without giving them the passwords. This mechanism is used by companies such as Amazon, Google, Facebook, Microsoft and Twitter to permit the users to share information about their accounts with third party applications or websites. Generally, OAuth provides clients a "secure delegated access" to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without providing credentials. Designed specifically to work with Hypertext Transfer Protocol (HTTP), OAuth essentially allows access tokens to be issued to third-party clients by an authorization server, with the approval of the resource owner. The third party then uses the access token to access the protected resources hosted by the resource server.

JSON Web Token is an Internet standard for creating data with optional signature and/or optional encryption whose payload holds JSON that asserts some number of claims. The tokens are signed either using a private secret or a public/private key. For example, a server could generate a token that has the claim "logged in as admin" and provide that to a client. The client could then use that token to prove

that it is logged in as admin. The tokens can be signed by one party's private key (usually the server's) so that party can subsequently verify the token is legitimate. If the other party, by some suitable and trustworthy means, is in possession of the corresponding public key, they too are able to verify the token's legitimacy. The tokens are designed to be compact, URL-safe, and usable especially in a web-browser Single-Sign-On (SSO) context. JWT claims can typically be used to pass identity of authenticated users between an identity provider and a service provider, or any other type of claims as required by business processes.

### D. API Gateway

The API Gateway is an API management tool which sits between a client and a collection of backend services, which serves as a reverse-proxy accepting all API calls, aggregating services (if necessary) and returns the obtained result. The necessity of the API Gateway for the project originated from two aspects:

- A way to share data assets via APIs, as well as discovering these APIs themselves was required.
- General services such as analytics, third parties offering their own services as part of the platform for added value can be integrated into the platform via the API gateway.

The API Gateway is a backend service which can also be handled via the platform's user interface for convenience. When a new API is added to the gateway, the API is defined as a new endpoint, requiring the user to provide the URL of the API (parametric URLs for REST APIs are fully supported as well), the method of the API (POST, GET, DELETE, etc.), a small description, the authentication method the system needs to use to call the API.

### E. Anonymization Framework

The Anonymisation component is responsible to implement the pseudonymisation and anonymisation of the platform data. The component includes the following sub-components:

- **Consent database:** A database which stores the data subjects who have provided consent to the system.
- **Framework database:** A database which contains the PIIs (Personally Identifiable Information) of all the data subjects.
- **Re-identification database:** A database which contains the original data of the data subjects or other data which can be used to match the pseudonymised (or anonymised) data to the data subjects. These data need to be pseudonymised (or anonymised) and their access is restricted only to the authorised personnel.
- **Exposed database:** A database which contains the pseudonymised data which are accessed and disseminated to the various parties which use the system.
- **Pseudonymisation:** A component which will perform pseudonymisation transformations on the data.
- **Anonymisation:** A component which will anonymise the data.
- **Data adapter:** A software component which is responsible to implement the pseudonymisation of the data.

## V. Conclusions

This work proposed a semantic data platform that aims to deliver sophisticated backbone service capabilities that will enable trusted, secure, automated, robust and controlled data transactions for food certification that aims to bring competitive advantages to all food sector businesses that demand easy, fast, and actionable access to variegating food safety data from multiple devices and in various settings. Our system is based on a a virtual environment that facilitates the exchange and connection of data between different organizations, through a shared reference architecture and common governance rules and enables trusted and secure sharing of food safety data assets. As future enhancements, authors of this work would like to explore other methods for supporting the facilitation of data sharing, such as machine learning training procedures and machine learning models for data from differing sources.

## Acknowledgment

This work is a part of TheFSM project, that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 871703.

## References

[1] W. Schlenker and S. B. Villas-Boas, "Consumer and market responses to mad cow disease," *American Journal of Agricultural Economics*, vol. 91, no. 4, pp. 1140–1152, 2009.

[2] J. Premanandh, "Horse meat scandal–a wake-up call for regulatory authorities," *Food control*, vol. 34, no. 2, pp. 568–569, 2013.

[3] gfsi, "The coalition of action on food safety." [Online]. Available: https://mygfsi.com/

[4] H. Bremmers and K. Purnhagen, "Regulating and managing food safety in the eu: a legal-economic perspective," in *Regulating and Managing Food Safety in the EU*. Springer, 2018, pp. 1–9.

[5] FDA, "Tracking and tracing of food." [Online]. Available: https://www.fda.gov/food/new-era-smarter-food-safety/tracking-and-tracing-food

[6] K. WANG, Z. CHEN, and J. XU, "Efficient traceability system for quality and safety of agricultural products based on consortium blockchain," *Journal of Computer Applications*, vol. 39, no. 8, p. 2438, 2019.

[7] M. Wysel, D. Baker, and W. Billingsley, "Data sharing platforms: How value is created from agricultural data," *Agricultural Systems*, vol. 193, p. 103241, 2021.

[8] B. Lang, D. M. Conroy *et al.*, "When food governance matters to consumer food choice: Consumer perception of and preference for food quality certifications," *Appetite*, vol. 168, p. 105688, 2022.

[9] V. Urovi, V. Jaiman, A. Angerer, and M. Dumontier, "Luce: A blockchain-based data sharing platform for monitoring data license accountability and compliance," *arXiv preprint arXiv:2202.11646*, 2022.

[10] R. K. Lomotey, S. Kumi, and R. Deters, "Data trusts as a service: Providing a platform for multi-party data sharing," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100075, 2022.

[11] N. Jha, M. Trevisan, L. Vassio, M. Mellia, S. Traverso, A. Garcia-Recuero, N. Laoutaris, A. Mehrjoo, S. A. Azcoitia, R. C. Rumin *et al.*, "A pims development kit for new personal data platforms," *IEEE Internet Computing*, vol. 26, no. 3, pp. 79–84, 2022.

[12] info@kraken.com. Kraken: the all-in-one fully integrated machine for multi-material hybrid manufacturing. [Online]. Available: https://krakenproject.eu/

[13] S. C. I. . T. de Informática (Spain). Dataports - a data platform for the connection of cognitive ports. [Online]. Available: https://dataports-project.eu/

[14] i3 market project. I3-market backplane - i3 market. [Online]. Available: https://www.i3-market.eu/i3-market-backplane/

[15] T. S. D. S. Space. Trusted secure data sharing space. [Online]. Available: https://www.trusts-data.eu/
[16] H. Feng, X. Wang, Y. Duan, J. Zhang, and X. Zhang, "Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges," *Journal of cleaner production*, vol. 260, p. 121031, 2020.
[17] I. González-Puetate, C. Marín-Tello, and H. R. Pineda, "Agri-food safety optimized by blockchain technology," *Revista Facultad Nacional de Agronomía Medellín*, vol. 75, no. 1, pp. 9839–9851, 2022.
[18] B. J. H. ter, *Semantic Data Modeling*. Prentice Hall, 1992.
[19] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," in *Annual international conference on the theory and applications of cryptographic techniques*. Springer, 2011, pp. 568–588.
[20] M. Cohn, *Agile softwareentwicklung MIT scrum zum erfolg!* Addison-Wesley, 2010.
[21] V. C. Hu, D. Ferraiolo, R. Kuhn, A. R. Friedman, A. J. Lang, M. M. Cogdell, A. Schnitzer, K. Sandlin, R. Miller, K. Scarfone *et al.*, "Guide to attribute based access control (abac) definition and considerations (draft)," *NIST special publication*, vol. 800, no. 162, pp. 1–54, 2013.