automation

MDPI

Article

The AgriTrust Framework: Federated Semantic Governance for Trusted and Interoperable Agricultural Data Sharing

Ivan Bergier [1,*], Jayme Garcia Arnal Barbedo [1], Édson Luis Bolfe [1], Debora Drucker [1] and Filipi Miranda Soares [2]

[1] Brazilian Agricultural Research Corporation, Embrapa Digital Agriculture, Campinas 13083-886, SP, Brazil; jayme.barbedo@embrapa.br (J.G.A.B.); edson.bolfe@embrapa.br (É.L.B.); debora.drucker@embrapa.br (D.D.)
[2] Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie (MISTEA), University of Montpellier, INRAE & Institut Agro, 2 Place Pierre Viala, 34060 Montpellier, France; filipi.miranda-soares@inrae.fr
* Correspondence: ivan.bergier@embrapa.br

Highlights

- Presents the AgriTrust Ontology for semantic agricultural data governance.
- Simulated across three Brazilian supply chains via a federated knowledge graph.
- Integrates governance (data contracts) with supply chain traceability (OWL).
- Minimizes the AgData Paradox: distrust and fragmentation hinder data value.

Abstract

New regulations, such as the EU Deforestation-Free Regulation (EUDR), make verifiable agricultural data (AgData) essential for global trade. However, its value is compromised by a widespread "AgData Paradox", characterized by distrust and fragmentation. To address this problem, we present AgriTrust, a federated semantic governance framework that automates and governs data sharing. Its key methodological innovation lies in the deep integration of a multi-sectorial governance model with a semantic digital layer, implemented through the AgriTrust Ontology (an OWL ontology for tokenization and traceability) and a multi-vendor, blockchain-agnostic architecture that avoids single-vendor dependence. We demonstrate the framework's feasibility through simulated case studies in three critical Brazilian supply chains: coffee (EUDR compliance), soybean (mass balance), and beef (animal traceability). Using a semantic reasoning pipeline on a proof-of-concept federated knowledge graph of 2010 triples, we show how AgriTrust enables verifiable provenance representation, automated compliance checking via executable data contracts, and cross-platform asset management. The results provide initial evidence that AgriTrust offers a conceptually coherent blueprint for agricultural data sharing, though operational deployment, scalability testing, and performance validation under real-world conditions remain as future work.

Keywords: agricultural data economy; data sovereignty; ODRL; semantic interoperability; SHACL; supply chain traceability; W3C Verifiable Credentials

Academic Editors: Ata Jahangir Moshayedi, Zeashan H. Khan and Amin Kolahdooz

Received: 13 February 2026
Revised: 13 March 2026
Accepted: 27 March 2026
Published: 31 March 2026

1. Introduction

Brazilian agribusiness is a cornerstone of the global food system and a frontier of technological adoption. The sector's digital transformation generates vast volumes of agricultural data (AgData) with immense potential to drive unprecedented gains in efficiency,

sustainability, and value creation [1,2]. This potential has become increasingly urgent due to regulations such as the EU Deforestation-Free Regulation (EUDR)—a European Union mandate requiring importers to verify that products are deforestation-free—which make verifiable AgData a concrete prerequisite for market access [3].

However, this opportunity is stifled by a pervasive "AgData Paradox" [4–6]. Formally, the paradox is the observed contradiction in the agricultural data economy: stakeholders universally recognize the high potential value of data sharing, yet this potential remains systematically unrealized due to three interrelated barriers. First, a trust deficit stems from producers' fears over data sovereignty and unfair value appropriation. Second, a lack of technical interoperability results from heterogeneous data sources and formats. Third, an absence of standardized, fair governance leads to unclear rules on data ownership, consent, and benefit sharing. This paradox manifests as fragmented data silos that directly impede critical outcomes such as verifiable supply chain traceability and automated regulatory compliance.

Current technological solutions often address only part of the problem. Isolated traceability platforms or singular blockchain implementations can ensure data integrity but often fail to address foundational governance challenges [4,7–10]: Who owns the data? Under what conditions can it be used? How are the benefits shared? Without a trusted, standardized, and fair governance framework, these solutions struggle to achieve widespread adoption and, crucially, interoperability across the ecosystem, risking the creation of new, more technologically advanced silos.

This limitation highlights a significant research gap at the intersection of data governance, semantic technologies, and distributed systems. While substantial work exists in each field independently, from the FAIR principles (Findable, Accessible, Interoperable, and Reusable)—a machine-oriented framework for data management—and CARE principles (Collective Benefit, Authority to Control, Responsibility, and Ethics)—a people-centric complement emphasizing Indigenous data sovereignty—to OWL (Ontology Web Language) and blockchain [11–14], their integration into a cohesive, operational framework for agriculture remains nascent. The critical missing piece is a foundational layer that seamlessly intertwines enforceable governance rules with machine-interpretable semantics.

This work builds directly on our research group's prior investigations into agricultural data governance and traceability. Bergier et al. [6] examined data reporting practices in agri-food platforms, identifying the critical need for producer-centric governance mechanisms. Concurrently, our development of the API BovTrace system (https://www.portal.agroapi.cnptia.embrapa.br/api-docs/bovtrace, accessed on 13 March 2026) for tracing cattle movements across Brazilian farms revealed practical interoperability challenges that could not be resolved through conventional platform integration alone. These prior efforts established the problem context and practical requirements that motivated the AgriTrust framework's design, particularly the emphasis on data sovereignty (ownedBy) and cross-platform interoperability (CrossChainToken) as first-class ontological constructs [7].

To bridge this gap, this work aims to present an enhanced version of AgriTrust, a federated semantic governance framework for trusted agricultural data sharing. The contributions are distinct from previous efforts in how they collectively address the holistic nature of data sharing by balancing technical, governance, and business concerns, where prior strategies have often addressed these concerns in isolation. More explicitly, its contributions are

1. An integrated governance-digital framework. We move beyond purely technical or abstract governance models by explicitly coupling a principled, multi-stakeholder governance model, built on data sovereignty, transparent data contracts, equitable

value sharing, and regulatory compliance, with a machine-interpretable semantic layer. This ensures that technical execution is inherently guided by and enforces fair business rules and stakeholder equity from the outset.

2. The AgriTrust Ontology. We provide a comprehensive, formal OWL ontology, built on established foundations in knowledge graph engineering [15], that creates a shared vocabulary for agricultural tokenization, traceability, and certification. Critically, it offers a machine-reasonable foundation for semantic interoperability, enabling not only asset tracking but also the complex logic required for automated compliance and certification via Shapes Constraint Language (SHACL) [16]. Its design includes native alignment with the pivotal W3C standards Verifiable Credentials [17], Open Digital Rights Language (ODRL) [18], and Data Catalog Vocabulary (DCAT-3) [19], ensuring compatibility within broader data ecosystems.
3. A blockchain-agnostic, multi-provider architecture. The proposed federated architecture directly counters a major adoption barrier, vendor lock-in, by decoupling the governance and semantic layers from any single ledger implementation. This design fosters a competitive, multi-provider ecosystem and provides a dual-layer model that uses blockchain both as an immutable provenance ledger and as a settlement layer for value sharing, transforming data sharing into a concretely incentivized activity.
4. Proof-of-concept validation via multi-commodity case studies and analytics. We demonstrate the framework's practical viability and adaptability through detailed instantiations across three critical Brazilian supply chains (coffee, soybean, and beef). Validation is conducted via a unified semantic reasoning and analytics pipeline, which processes an instantiated knowledge graph, materializes inferred knowledge, and executes federated queries. This provides quantitative evidence that the framework successfully enables verifiable provenance, automates compliance, and manages assets across multiple platforms, moving from theoretical potential to validated, operational utility.

Together, these contributions make AgriTrust a practical and sustainable blueprint for scalable, trusted agricultural data sharing toward sustainable food production and consumption.

The remainder of this paper is organized as follows. Section 2 reviews the relevant background and related work. Section 3 details the AgriTrust framework, encompassing its governance model, core ontology, and technical architecture. Section 4 presents the implementation and validation, detailing the ontology instantiation for multi-commodity case studies and the results from a dedicated analytical pipeline. Section 5 discusses AgriTrust security; privacy and resilience; economic, environmental, and social impacts; comparisons of the framework with existing approaches; and limitations and challenges for future work. The final section highlights the main conclusions.

2. Background and Related Work

The AgriTrust framework is positioned at the confluence of several mature research streams: data governance, semantic technologies, tokenization, and federated architectures. This section reviews these domains, highlighting both their individual contributions and the critical gaps that emerge at their intersection, gaps that AgriTrust is designed to fill through principled synthesis.

2.1. Data Governance and Sovereignty in Agriculture

General data governance frameworks emphasize transparency, accountability, and fairness [9,13]. However, agriculture demands models that account for the producer's unique position. The concept of data sovereignty, i.e., the right of data originators to

govern data from their land, is paramount [4,13]. This is powerfully articulated by the CARE Principles for Indigenous Data Governance, which serve as a crucial people-centric complement to the well-known machine-oriented FAIR principles [12,13].

While these principles provide an essential ethical and conceptual foundation, a significant gap exists in their technical instantiation. How is "Authority to Control" enforced in a multi-platform digital ecosystem? AgriTrust addresses this by embedding producer sovereignty as a non-negotiable, technically enforced pillar within its operational framework, directly tackling the trust deficit identified in private AgData ecosystems [20,21].

2.2. Semantic Technologies and Interoperability

Overcoming syntactic and semantic heterogeneity is a prerequisite for data sharing. Ontologies, formalized using standards like the OWL, provide a shared, machine-interpretable vocabulary for a domain [11]. In agriculture, significant community efforts have established semantic foundations: repositories like AgroPortal curate domain ontologies [22–24], while initiatives like the Agricultural Data Exchange (ADE) and ISO/TC 347 work on standardized data models [25].

Building on this foundation, recent work advances semantic interoperability for agricultural data. Soares [24] proposes an integrated framework combining ontologies, knowledge graphs, and AI to resolve data conflicts in existing, heterogeneous datasets, a critical step for retroactive harmonization. Parallel work explores the use of semantic resources with explainable AI to address trust and integration in smart farming [23].

This landscape reveals a bifurcation: one strand focuses on retrospective harmonization of legacy data [23,24], while this work focuses on prospective orchestration of new data creation and sharing within a governed ecosystem. AgriTrust contributes a dedicated ontology that provides not just domain description but the precise semantic machinery for tokenization and machine-executable governance, enabling shared operational understanding across platforms.

2.3. Tokenization, Blockchain, and Digital Assets

Tokenization creates a cryptographic digital representation of an asset on a ledger, enabling traceability. Research has extensively explored blockchain for supply chain integrity, from agri-food to pharmaceuticals [26,27]. However, many implementations remain siloed, relying on proprietary models and single-chain architectures that inhibit broader ecosystem integration and replicate data fragmentation at a higher technological tier [5,9,23,24].

These systems often prioritize immutability but lack granular, legally enforceable models for access control and value sharing. Contemporary infrastructures like AgriFoodTEF (https://dataspace.agrifoodtef.eu, accessed on 13 March 2026) and AgrospaAI (https://portal.agrospai.udl.cat, accessed on 13 March 2026), built on the Pontus-X technology (https://www.pontus-x.eu, accessed on 13 March 2026), represent an evolution by integrating access control and policy enforcement directly into a federated architecture for data and service exchange.

AgriTrust builds on this paradigm by explicitly formalizing the semantic layer required for Tokenization-as-a-Service (TaaS) interoperability. Rather than claiming novelty in the TaaS concept itself, the framework contributes the ontological machinery, specifically the CrossChainToken class and blockchain-agnostic properties, that enables diverse TaaS platforms to interoperate semantically while preserving data sovereignty. This elevates tokenization from a siloed technical mechanism to a domain-level service explicitly linked to certification semantics and transparent governance, thereby enhancing the trustworthiness of digital assets across platform boundaries.

2.4. Data Spaces and Federated Architectures

The vision of data spaces, exemplified by the International Data Spaces (IDS) reference architecture, provides the most aligned blueprint for sovereign, federated data sharing [28]. A data space is a governed infrastructure connecting distributed sources via common rules, where participants retain control, a philosophy central to AgriTrust.

The Common European Agricultural Data Space (CEADS) embodies this vision for the agricultural sector, emphasizing interoperability-by-design, semantic alignment, and compliance with European data strategies [29,30]. It promotes a federated model where multiple infrastructures interoperate through shared standards, APIs, and governance, rather than a single centralized platform.

2.5. Synthesis and Identified Gap

The related work reveals significant yet disconnected progress. We have ethical governance principles (CARE/FAIR), technical semantic tools (ontologies), immutable ledgers (blockchain), and federated architectural blueprints (IDS/CEADS). The critical gap is not a lack of components, but the absence of a domain-specific framework that deeply integrates them into an operational whole. Existing ontologies often lack governance artifacts; governance models remain abstract; blockchain solutions create new silos; and data space architectures require concrete, sector-specific implementation.

To crystallize the research gap, Table 1 maps the three components of the AgData Paradox (trust deficit, interoperability barriers, and governance absence) against the four approach categories reviewed above. Isolated blockchain platforms address trust through immutability but create new silos that exacerbate interoperability challenges and lack governance models for data sovereignty [8,9]. Centralized platforms offer governance within their boundaries but concentrate control, undermining producer trust and creating dependency [4,22]. Semantic ontologies resolve interoperability through shared vocabularies but stop at descriptions, lacking executable governance mechanisms [11,24]. Data space architectures provide the correct governance principles and federation blueprint but require concrete sectoral instantiation with machine-interpretable semantics [28–30].

Table 1. Mapping approach categories to AgData Paradox components.

Approach Category	Isolated Blockchain Platforms	Centralized Platforms	Semantic Ontologies	Data Space Architectures
Trust Deficit Addressed?	✓ (immutability)	✗ (vendor lock-in)	✗ (descriptive only)	✓ (sovereignty principles)
Interoperability Addressed?	✗ (new silos)	✗ (proprietary formats)	✓ (shared vocabulary)	✓ (federation model)
Governance Addressed?	✗ (no sovereignty model)	✓ (within platform)	✗ (no execution)	✓ (policy framework)
Critical Gap	Cross-platform interoperability; producer control	Data portability; fair value sharing	Executable governance; tokenization	Concrete agricultural semantics; instantiation

The critical gap revealed by this mapping is not a lack of components, but the absence of a domain-specific framework that deeply integrates them into an operational whole. AgriTrust is designed to fill this integration gap. However, to understand precisely how AgriTrust differs from existing approaches, it is necessary to examine specific initiatives that informed (and are extended by) the framework.

AgroPortal [22] provides an essential infrastructure for hosting and curating agricultural ontologies, including the AgriTrust Ontology itself (now available at https://w3id.org/agrisemantics/AGRITRUST, accessed on 13 March 2026). AgroPortal's role as a registry and

its support for ontology versioning, alignment, and community feedback were instrumental in ensuring AgriTrust's FAIR compliance. However, AgroPortal is primarily a repository for ontologies, not a framework for executable governance or cross-platform data sharing. AgriTrust follows a similar approach by embedding governance artifacts (data contracts and certificates) directly into the ontology and linking them to executable standards (ODRL and W3C VC), thereby transforming static semantic models into operational components of a data sharing ecosystem.

The work of Soares [24] proposes a semantic interoperability framework for retroactive harmonization of heterogeneous agricultural datasets, a critical contribution for integrating legacy data. This work directly informed AgriTrust's approach to semantic alignment, particularly the reuse of foundational ontologies (PROV-O and SOSA/SSN) and the adoption of knowledge graph principles [15]. However, Soares' framework focuses on reconciling existing data after the fact, whereas AgriTrust addresses prospective governance and interoperability by design. Where Soares provides mechanisms to ask "how can we make these disparate datasets speak to each other?", AgriTrust asks "how can we ensure that new data is created and shared within a governed ecosystem from the outset?" The two approaches are complementary: AgriTrust's semantic layer ensures that data created within the framework is natively interoperable, while Soares' methods could be applied to integrate legacy data into the same ecosystem.

CEADS and IDS [28–30] articulate the high-level vision and architectural principles for sovereign, federated data spaces. AgriTrust draws directly from these blueprints: the emphasis on data sovereignty, the federated multi-provider model, and the policy-based access control all align with IDS and CEADS specifications. However, these initiatives provide reference architectures, not concrete, sector-specific implementations. The critical challenge they leave open is: What does a data space look like for agriculture at the operational level? AgriTrust addresses this gap by providing (a) a domain ontology that explicitly models agricultural concepts (assets, batches, and processes) alongside governance artifacts (contracts, certificates, and tokens); (b) machine-executable semantics via ODRL and SHACL, moving from policy descriptions to enforceable rules; (c) multi-provider extensibility patterns that allow diverse platforms to interoperate without centralized control; and (d) concrete tokenization semantics that link physical assets to digital representations across blockchains.

From this analysis, four particular, distinctive elements emerge that collectively constitute the AgriTrust differential:

1. Data sovereignty as a first-class ontological property (ownedBy). While data sovereignty is discussed in governance frameworks (CARE principles) and data space architectures (IDSs), existing approaches treat it as a policy principle rather than a machine-enforceable constraint. In AgriTrust, every :Asset is linked to its :DataProducer via the :ownedBy property at the moment of instantiation, making sovereignty technically non-negotiable and auditable via SPARQL queries and SHACL validation. No other framework encodes ownership as a formal, queryable, and constrainable property within the asset model itself.
2. Cross-chain tokenization semantics (CrossChainToken and hasCrossChainReference). Existing blockchain platforms operate within single ledgers, creating new silos that mirror the fragmentation they aim to solve. AgriTrust introduces ontological constructs that explicitly model assets recognized across multiple blockchains, enabling semantic interoperability without requiring cross-chain smart contract execution. The :CrossChainToken class and :hasCrossChainReference property allow assets to be registered on different ledgers (e.g., Hyperledger for domestic traceability and Polygon

for international certification) while maintaining a unified semantic identity (absent in single-ledger implementations).

3. Governance artifacts as executable code (DataContract linked to ODRL policies, and certificates as W3C VC). While ODRL and verifiable credentials exist as standalone W3C standards, they are typically implemented as external layers bolted onto data models. AgriTrust integrates them directly into the domain ontology: :DataContract is linked to :ODRLPolicy via :hasODRLPolicy, and :Certificate is defined as a subclass of vc:VerifiableCredential. This integration transforms abstract governance principles into machine-executable rules that can be validated (via SHACL), queried (via SPARQL), and enforced (via smart contracts or API gateways) as an inherent part of the data graph.
4. Multi-provider extensibility by design. Most ontologies are designed as closed, monolithic artifacts. AgriTrust is engineered for federated evolution: the core ontology is kept minimal and stable, while platform providers can define domain-specific subclasses (e.g., beef:FinishingOperation and soybean:YieldPerHectare) without breaking interoperability. This enables competition and specialization. For instance, a herd management platform and a food safety platform can interoperate using the same core semantics while extending them for their specific needs. This pattern, encoded in the ontology's design, is absent in both monolithic platforms (which resist extension) and single-ontology approaches (which resist variation).

Together, these four elements constitute the AgriTrust differential: a framework that neither describes data nor governs access but operationalizes trustworthy, cross-platform data sharing through deep semantic integration of sovereignty, tokenization, and executable governance. While existing approaches address parts of the problem, AgriTrust is the first to weave them into a coherent, machine-interpretable, and operationally executable whole.

3. The Framework: Integrated Governance and Semantic Architecture

This section presents the AgriTrust framework, an integrated model designed to resolve the AgData Paradox by intertwining a robust multi-stakeholder governance framework with a semantic digital layer. The core innovation lies in ensuring that technical interoperability is not an afterthought but is guided and enforced by clear, fair, and transparent rules of engagement from the outset.

3.1. Framework Overview and Design Principles

The AgriTrust framework was architected as a federated ecosystem rather than a centralized platform. In this process, generative AI (DeepSeek-V3-0324, with earlier contributions from ChatGPT-4o) assisted in exploring OWL ontology patterns, generating Turtle syntax for core classes, formalizing governance artifacts (data contracts, ODRL policies, and W3C VCs), and synthesizing the literature into a coherent structure.

A federated model refers to an architectural paradigm that connects a distributed network of independent, autonomous nodes (e.g., Token-as-a-Service or TaaS platforms, and data sources) through a common set of governance rules and technical standards. Unlike a centralized system that consolidates data and control into a single entity, a federated system preserves sovereignty (each participant retains control over their own data and infrastructure), ensures interoperability (a shared semantic layer or the core ontology and standardized interfaces enable seamless communication and data exchange between nodes), and prevents monopoly (the ecosystem is designed for multi-provider participation, avoiding vendor lock-in and fostering innovation through competition).

This model creates a cohesive data sharing environment without requiring data to be physically centralized, thereby directly upholding the principle of data sovereignty. It connects distributed data sources, managed by various platform providers, through a common set of governance rules and semantic standards. The high-level architecture is depicted in Figure 1.

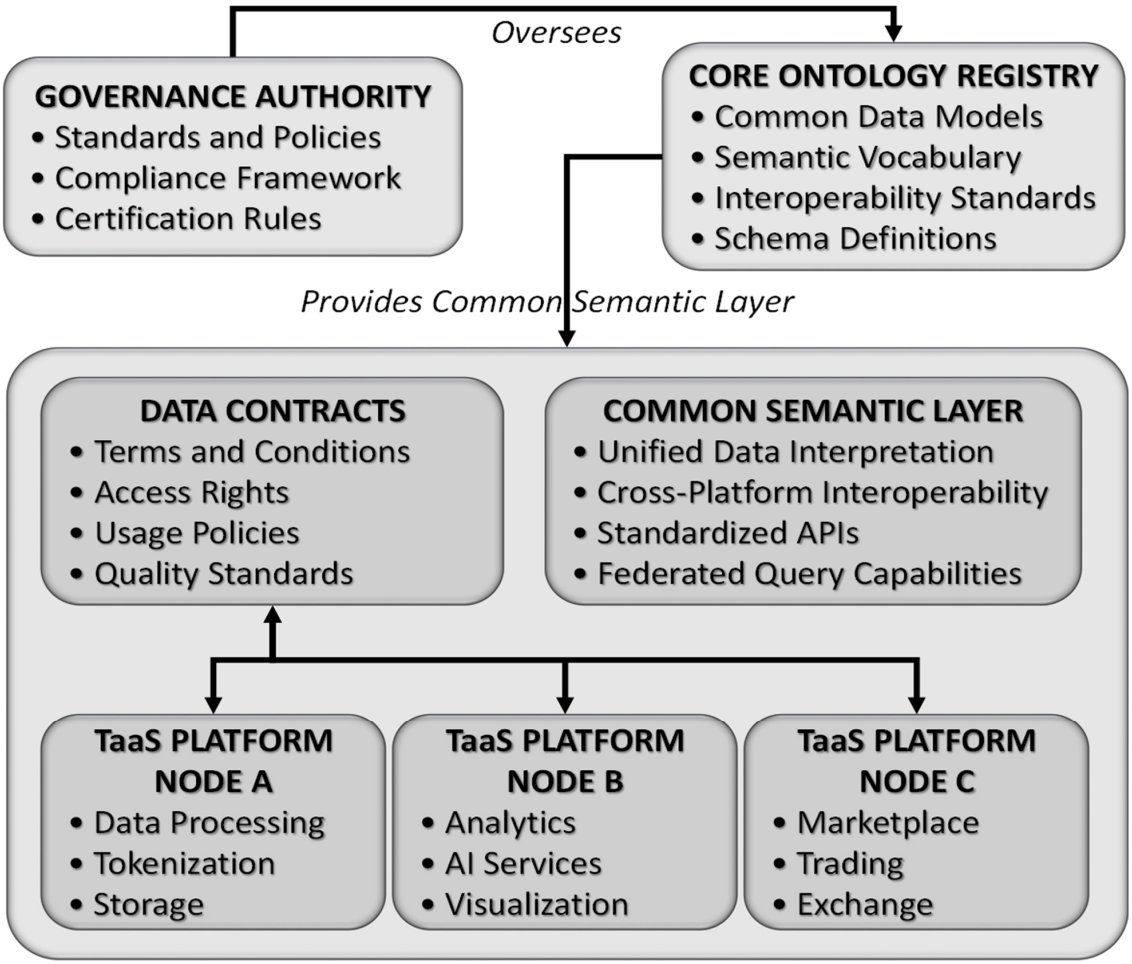

Figure 1. Overview of the integrated governance and digital AgData sharing ecosystem, showing governance authority, core ontology registry, and multiple TaaS platform nodes (A, B, and C) interacting via data contracts in a Common Semantic Layer.

The framework's design is architected around four core principles:

1. Data sovereignty as the non-negotiable foundation. Producers and data originators retain ultimate ownership and control. The technical system is engineered to enforce this authority by design, ensuring producers govern how their data is accessed, used, and shared, thereby addressing the fundamental trust deficit;
2. Governance-led technical implementation. The multi-stakeholder governance model defines the explicit "rules of the game". The digital infrastructure, the ontology, APIs, and smart contracts are designed not as neutral tools but as a faithful execution layer for these rules, making compliant and equitable operation the systemic default;
3. A federated, not centralized, architecture. The ecosystem is predicated on the participation of multiple, independent platform and blockchain providers. This structure intentionally avoids monopolistic control and vendor lock-in, instead fostering innovation, resilience, and specialization through competition within a standardized framework;

4. Semantic interoperability as the core enabler. Recognizing that syntactic data exchange is insufficient for trust, a shared, unambiguous understanding of meaning is established as a prerequisite. The formal AgriTrust Ontology provides this common language, enabling both humans and machines to interpret concepts like "sustainable batch" or "data contract" consistently across all platform boundaries.

3.2. Governance Model

The governance model establishes the rules, roles, and processes that create a fair, transparent, and enforceable environment for all participants.

3.2.1. Multi-Stakeholder Governance Authority

A cornerstone of AgriTrust is the governance authority, an entity (typically a consortium comprising producer associations, industry representatives, certifiers, and government bodies) responsible for the ecosystem's integrity. Its mandates include

- Maintaining and evolving the core governance framework;
- Managing the version-controlled core ontology registry;
- Accrediting platform and blockchain providers;
- Providing a dispute resolution mechanism.

3.2.2. Core Governance Pillars

The framework's stability is built upon four pillars, instantiated through legal and technological instruments:

1. Producers and data originators retain ultimate authority, technically enforced via access control mechanisms in data contracts;
2. Machine-readable data contracts (detailed in Section 3.4.1) govern all data sharing transactions. These are digitally signed agreements that specify the purpose, duration, scope, and terms of data use, creating an auditable trail of all data-related transactions;
3. The model explicitly recognizes that data has economic value and promotes mechanisms for the fair distribution of benefits derived from shared data among all contributors. This can be operationalized through automated premium payments or access-to-service agreements encoded within data contracts;
4. The framework is designed to ensure that data sharing practices inherently comply with relevant laws and norms, such as Brazil's General Data Protection Law (LGPD), which establishes rules for collecting and processing personal data), and international standards like the EU Deforestation-Free Regulation (EUDR).

3.2.3. Roles and Responsibilities

Clear roles delineate responsibilities and expectations within the ecosystem:

- Data producer: the entity from which data originates (e.g., a farmer or a cooperative providing yield, sensor, or management data); they are the sovereign owners of their data;
- Data consumer: the entity that uses data under specific terms defined in a data contract (e.g., a certifier validating sustainable practices and a market requesting proof of origin);
- Platform provider: the entity that offers the technical infrastructure (e.g., a TaaS platform) for creating and managing digital tokens and enforcing data contracts; they must maintain neutrality and adhere to the governance rules;
- Certifier: a trusted third-party entity responsible for auditing and issuing compliance certificates (e.g., sustainable and deforestation-free) based on verifiable data accessed via data contracts.

3.3. Semantic Layer: The AgriTrust Ontology

The AgriTrust Ontology is a formal OWL 2 ontology engineered to create a shared conceptual foundation for tokenization, traceability, and governed data sharing in agricultural supply chains. It addresses the semantic layer of the "AgData Paradox" by providing unambiguous, machine-interpretable definitions for assets, agents, processes, and the governance artifacts that connect them.

3.3.1. Design Principles and Top-Level Structure

The ontology was engineered with four core principles:

1. Data sovereignty by design: the :ownedBy property is a first-class relationship, explicitly linking every :Asset to its owning :DataProducer from the moment of instantiation;
2. Governance as part of the data model: governance artifacts like :DataContract and :Certificate are integral classes, linked to assets and tokens via formal properties (:governedBy and :hasCertificate), ensuring rules are part of the data graph;
3. Blockchain agnosticism: the :BlockchainProvider class and the :registeredOnBlockchain property decouple logical assets and tokens from any specific ledger implementation, while :CrossChainReference enables multi-provider interoperability;
4. Process-centric provenance: the :Process class and properties (:hasInput, :hasOutput, and :hasProvenance) explicitly model transformations, moving beyond simple location tracking to capture the "story" of an asset.

As shown in the diagram below (Figure 2), the ontology structures the domain into four interconnected blocks: assets and tokenization, agents and governance, process and provenance, and observations and metrics. This structure ensures that a digital token is not an isolated record but a node in a rich graph of ownership, transformation, and agreed-upon rules.

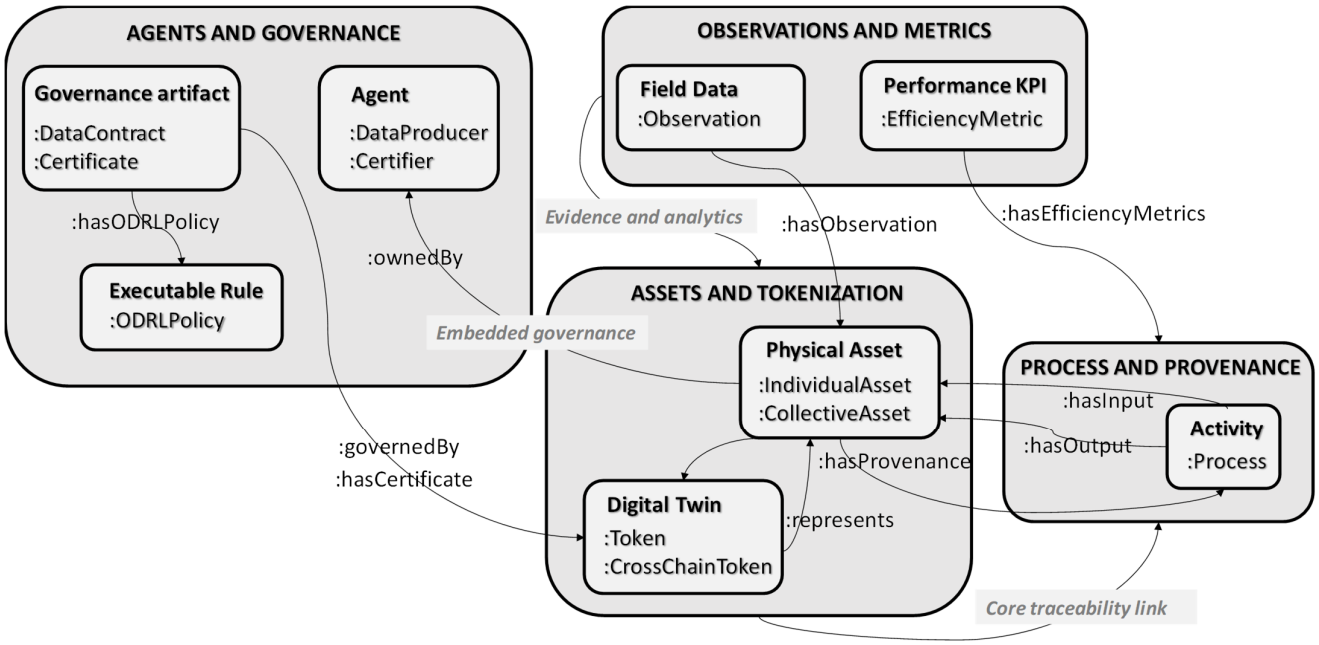

Figure 2. Conceptual model of the AgriTrust Ontology, showing the four interconnected pillars that enable semantic interoperability.

Another feature is that the AgriTrust Ontology reuses and aligns with well-established standards to ensure broad interoperability:

- PROV-O for modeling provenance, ensuring a clear chain of custody for all assets (https://www.w3.org/TR/prov-o/, accessed on 13 March 2026);
- SOSA/SSN for representing observations and measurements from sensors and manual inputs (https://www.w3.org/TR/vocab-ssn-2023/, accessed on 13 March 2026);
- GeoSPARQL for representing spatial geometries of farms and plots (http://www.opengis.net/ont/geosparql, accessed on 13 March 2026).

To ensure native interoperability within modern data spaces, the ontology extends this foundation with explicit mappings to three technical standards:

- W3C Verifiable Credentials (VCs) to make its :Certificate class a cryptographically verifiable digital attestation [17];
- Open Digital Rights Language (ODRL) to provide a machine-executable governance layer for its :DataContract class, enabling automated policy enforcement [18];
- DCAT-3 to ensure its :Dataset, :Agent, and other core resources are discoverable within standardized data catalogs and marketplaces [19].

3.3.2. Core Classes and Properties

The ontology defines 25 core classes and 30 object/datatype properties. Key elements include the following.

Core Traceability (:Asset, :Token, and :Process)

The :represents property forms the fundamental link between a physical asset and its digital token. Processes are explicitly connected to their inputs and outputs, enabling graph-based provenance queries (e.g., "What processes transformed this coffee batch?").

Governance (:DataContract, :Certificate, and :ODRLPolicy)

A :DataContract :coversAsset and :coversObservation, and is linked via :hasODRLPolicy to a machine-executable ODRL policy. A :Certificate is defined as a subclass of vc:VerifiableCredential, incorporating standards-compliant properties like :issuer (Decentralized Identifier or DID) and :proof.

Interoperability (:CrossChainToken and :BlockchainProvider)

The :CrossChainToken class and :hasCrossChainReference property model assets that are acknowledged across multiple ledgers. The :registeredOnBlockchain property allows any entity to declare its system of record without being bound to a single provider.

Data and Metrics (:Observation and :EfficiencyMetric)

These classes allow field data (e.g., sensor readings) and calculated performance indicators (e.g., water usage efficiency) to be attached to assets, processes, or agents, supporting both compliance and analytical use cases.

3.3.3. Formal Constraints and Automated Reasoning

To ensure data quality and consistency, the ontology is complemented by SHACL shapes [16]. These shapes enforce critical business rules, such as

- A :Token must :represent exactly one :Asset;
- A :DataContract must have a valid period (:validFrom and :validUntil) and at least one :hasPolicyAssignee;
- A :Certificate must have an issuer DID and an issuance date.

As shown further, these constraints are validated during data ingestion, guaranteeing that all instantiations conform to the framework's rules.

3.3.4. Multi-Provider Extensibility Pattern

A critical feature of the AgriTrust Ontology is its design for extensibility. The core is kept minimal and stable. Platform providers can define their own subclasses and properties for domain-specific concepts without breaking interoperability. For example, a beef platform can create a beef :FinishingOperation subclass of :Process, while a soybean platform can create soybean :YieldPerHectare as a subclass of :EfficiencyMetric. As long as all platforms commit to the core ontology, they retain the ability to query and understand the fundamental relationships across the ecosystem.

3.4. Technical Architecture

The integration of governance and semantics is realized through a federated technical architecture that embeds the rules directly into the data sharing infrastructure.

3.4.1. Core Components

The architecture of the ontology is composed of the following key components:

- Core ontology registry: a central, version-controlled repository for the shared OWL ontology, maintained by the governance authority (in this case, the AgroPortal at https://agroportal.lirmm.fr/ontologies/AGRITRUST, accessed on 13 March 2026). All platform providers must synchronize with this registry to ensure a consistent semantic understanding across the ecosystem;
- TaaS platforms: independent platforms operated by cooperatives, agroindustries, or third-party providers. They implement the core ontology and are responsible for
    - Tokenizing assets (:Asset → :Token);
    - Hosting and serving verifiable data about assets and processes;
    - Enforcing data contracts via smart contracts or secure, policy-enforced APIs;
    - Identity and Access Management (IAM): a decentralized system (e.g., based on DIDs) for authenticating all participants (agents and platforms) within the ecosystem;
    - Query interfaces: standardized APIs (primarily SPARQL endpoints) that allow participants to perform federated queries across different TaaS platforms using the common ontological terms.

The governance authority maintains the core ontology registry. Accredited TaaS platforms synchronize with this registry and provide services like tokenization and data hosting. Participants are authenticated via a decentralized IAM system. Data consumers (e.g., certifiers and retailers) interact with the ecosystem through standardized query interfaces to execute federated queries across the TaaS platforms, with all data access governed by machine-readable data contracts.

3.4.2. Data Contracts: The Operational Feature

The :DataContract is the pivotal entity that connects the governance and technical layers. It is a machine-readable manifestation of a legal agreement. Technologically, these contracts can be implemented as smart contracts on a blockchain or as policy documents in a secure API gateway. Their execution is automated and access is granted only when contract conditions are met and revoked upon expiration or breach.

For a more realistic perspective, consider a data contract for sustainable certification, which governs the sharing of specific farm data with a certification body, including clear usage restrictions and value sharing terms. Such a contract exemplifies how the four governance pillars (Section 3.1) are translated into executable code.

For greater clarity, Listing 1 provides a simplified Turtle representation of the DataContract_CoffeeExport instantiated in our validation dataset. This example illustrates how the four governance pillars are translated into machine-readable form: the :hasPolicyAssignee property identifies the authorized data consumer (sovereignty enforcement); the ODRL policy embedded via :hasODRLPolicy specifies the permitted action (odrl:read) and purpose ("EU import compliance") in executable form (transparent contract); the :coversAsset and :coversObservation properties link the contract to specific data resources (value sharing context); and the combination of purpose and validity period supports regulatory compliance verification.

Listing 1. Turtle representation of a data contract instance.

```
:DataContract_CoffeeExport a :DataContract ;
    :contractId "CONTRACT-EUDR-2024-001" ;
    :validFrom "2024-06-01T00:00:00Z"^^xsd:dateTime ;
    :validUntil "2025-06-01T00:00:00Z"^^xsd:dateTime ;
    :purpose "EU import compliance" ;
    :hasPolicyAssignee :Consumer_EUImporter ;
    :hasODRLPolicy :ODRLPolicy_CoffeeExport ;
    :coversAsset :CoffeeBatch_2024_001 ;
    :coversObservation :Observation_SoilPH .

:ODRLPolicy_CoffeeExport a odrl:Set ;
    odrl:permission [
        odrl:action odrl:read ;
        odrl:target :CoffeeBatch_2024_001
    ] ;
odrl:prohibition [
        odrl:action odrl:modify ;
        odrl:target :CoffeeBatch_2024_001
    ] .
```

Regarding ODRL integration, the framework adopts a direct import and instantiation approach rather than embedding policies as strings. The AgriTrust Ontology imports ODRL's core vocabulary (http://www.w3.org/ns/odrl/2/, accessed on 13 March 2026), allowing policies to be instantiated as first-class RDF nodes using ODRL classes and properties. As shown in Listing 1, :ODRLPolicy_CoffeeExport is typed as odrl:Set, with permissions and prohibitions structured according to the ODRL model. This enables policy reasoning and validation using standard ODRL tools while maintaining full integration with the AgriTrust graph. The :hasODRLPolicy property links the :DataContract to its executable policy, and the contract's :purpose provides human-readable context that aligns with the machine-readable ODRL permissions. For deployment, these ODRL policies could be compiled to platform-specific enforcement mechanisms (e.g., Hyperledger Fabric chaincode or XACML policies in an API gateway), though this translation step was not implemented in the current proof of concept.

Concerning enforcement mechanics, it is important to clarify the scope of the current validation. The Python 3.13 pipeline (Section 4.1.2) simulates contract enforcement through SHACL validation and ODRL policy checking at the semantic level: the pipeline verifies that contract instances conform to the ontology's structural constraints (e.g., required properties and valid date ranges) and that the associated ODRL policies are well-formed. However, integration with actual smart contract execution environments (e.g., deploying the ODRL policy as an Ethereum smart contract or Hyperledger Fabric chaincode) was not performed in this proof-of-concept study. Such integration represents a natural next step for operational deployment, where the semantic contract would be compiled to platform-specific executable code and enforcement would occur at both the API gateway (for off-chain data access) and blockchain layer (for on-chain state changes).

3.4.3. Blockchain-Agnostic Implementation

A key innovation of AgriTrust is its deliberate blockchain-agnosticism. The framework does not prescribe a single blockchain provider. Instead, the ontology includes properties for registering entities across various ledgers. This allows a token representing a Brazilian coffee batch to be registered on a national, low-cost blockchain (e.g., :BrazilAgriChain), while its EUDR compliance certificate is registered on a blockchain favored by European markets (e.g., :EUCertChain). The :hasCrossChainReference property can then link these entities, creating a verifiable, cross-chain provenance record without relying on a single, monolithic ledger.

It is crucial to define the precise scope of this blockchain-agnostic design. The AgriTrust framework is agnostic at the layer of asset registration and provenance anchoring. Its ontology provides the semantic layer to reference and link entities (tokens and certificates) across different ledgers. This successfully prevents vendor lock-in, allowing participants to select ledgers based on cost, jurisdiction, or technical requirements. However, this design does not inherently solve cross-chain logic execution as the challenge where a smart contract on one blockchain autonomously triggers a state change or payment on another. Orchestrating such atomic, cross-chain transactions requires specialized protocols and is a recognized research challenge in distributed systems. Therefore, AgriTrust's primary contribution is to establish semantic and referential interoperability for traceability and certification across chains. Solving atomic cross-chain transactions (e.g., a payment on Ethereum triggered by an event on Polygon) is outside the current scope and represents a key direction for future research.

4. Implementation and Multi-Commodity Validation

This section describes the implementation and empirical validation of the AgriTrust framework. We first detail the methodology and validation pipeline (see the Data Availability Statement section). We then present four detailed case studies instantiating the framework across structurally distinct supply chains. Finally, we synthesize the cross-commodity results to provide quantitative evidence of the framework's core capabilities.

4.1. Validation Methodology and Pipeline

It is important to clarify the scope and intent of the validation presented in this section. The primary goal of the current study is to establish and demonstrate the conceptual and semantic coherence of the AgriTrust framework, i.e., to verify that the ontology can represent the required domain concepts, that governance artifacts can be instantiated and linked, and that federated queries across simulated platforms return semantically meaningful results. Accordingly, the validation dataset was deliberately crafted as a proof-of-concept knowledge graph that exercises all core ontology classes and relationships across

structurally distinct supply chains. This approach follows established practices in ontology engineering [22,23].

Competency coverage and comprehensive scalability analysis, including stress tests on SPARQL query latency, OWL reasoning time, and SHACL validation throughput with datasets orders of magnitude larger, is reserved for complementary work currently underway. This companion research systematically evaluates AgriTrust's performance across synthetic datasets of increasing sizes, measuring reasoning times, query response curves, and memory utilization under load. Additionally, ongoing collaboration with industry partners can enable validation against real-world data inconsistencies and edge cases that cannot be anticipated in a hand-crafted dataset. The present paper thus establishes the functional correctness of the framework, while scalability and robustness are addressed in subsequent contributions.

4.1.1. Composition of the Validation Dataset

The dataset was designed to simulate a realistic, federated agricultural data ecosystem within the AgriTrust Ontology. It spans three high-value Brazilian supply chains and involves multiple independent actors to rigorously test interoperability and governance:

- Scope: Four physical assets (two coffee batches, one soybean batch, and one individual cattle), four digital tokens, three operational processes, two verifiable certificates, and two data contracts with linked ODRL policies;
- Multi-provider environment: Assets and tokens are deliberately distributed across three different simulated blockchain providers, Hyperledger Fabric (primary registry), Ethereum Mainnet, and Polygon, testing the blockchain-agnostic principle;
- Governance artifacts: These include a :DataContract for research data sharing (covering soy) and one for EU export compliance (covering coffee), each with a distinct purpose, validity period, and assignee;
- Data foundation: The integrated Turtle file (Ontology_instances.ttl) contains 627 explicit triples, the foundational, asserted facts that form our test knowledge base.

4.1.2. Analytical Pipeline Architecture

The validation was executed via a custom Python pipeline (available via GitLab; see Data Availability Statement). Its three-stage workflow is summarized in Figure 3 and designed to test the framework's core capabilities as follows:

- Load and Reason: The pipeline loads the Turtle file and applies OWL 2 RL reasoning using the owl-rl library. This materializes implicit knowledge based on the ontology's logical constraints, critically expanding the graph's inferential power;
- Query and Validate: A battery of SPARQL queries is executed against the reasoned graph. Each query module tests a specific framework capability: asset inventory, tokenization summary, blockchain distribution, process provenance, certificate verification, and data contract inspection;
- Generate Report: The pipeline aggregates entity counts into summary statistics and exports the complete reasoned graph, providing a reusable, logically consistent dataset.

The pipeline executes OWL 2 RL reasoning and SHACL validation sequentially: reasoning first materializes implicit knowledge (e.g., inferring that a :CrossChainToken is a :Token); then, SHACL validation checks the expanded graph against integrity constraints. This order ensures that shapes validate the complete inferred state rather than only asserted facts. In the case of conflict (for example, if reasoning infers a triple that violates a SHACL constraint), the validation step flags the violation, indicating inconsistency between the ontology's logical implications and the intended business rules. Such conflicts would signal modeling errors requiring ontology revision. In our validation dataset, no conflicts

occurred, indicating alignment between inferred semantics and governance constraints. This methodological approach provides a transparent, repeatable, and technically grounded validation of the AgriTrust operational semantics.

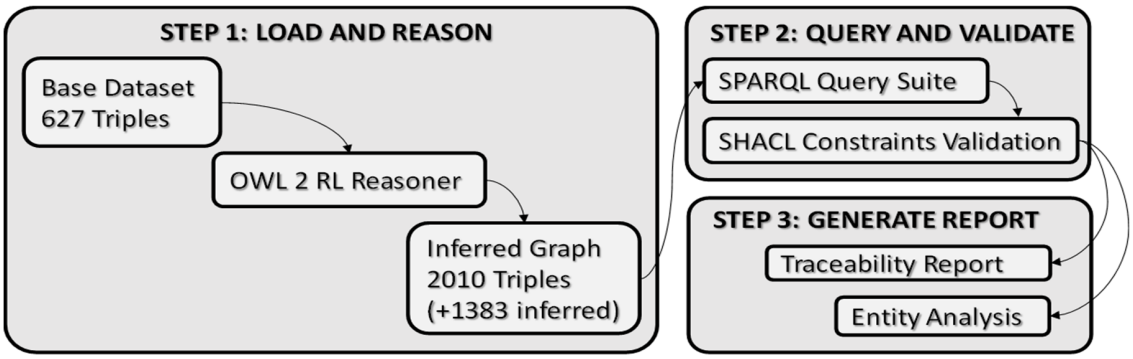

Figure 3. The three-step validation pipeline.

4.2. Case Study 1: Brazilian Coffee Supply Chain and EUDR Compliance

This case study applies AgriTrust to the coffee supply chain, where verifiable provenance and compliance with regulations like the EUDR are critical for market access.

4.2.1. Ontology Instantiation and Core Tokenization

We modeled a premium Arabica coffee batch as a :CollectiveAsset (CoffeeBatch_2024_001), explicitly linked to its owner (Farm_SantaMaria) via the :ownedBy property, embedding data sovereignty as a foundational triple (Table 2). Tokenization was demonstrated by creating a digital twin, Token_Coffee_001 (a :Token that :represents the physical batch). This token was :registeredOnBlockchain on Provider_Hyperledger and :managedByPlatform by the corresponding TaaS node, creating an immutable, platform-agnostic anchor for the asset's digital identity.

Table 2. Coffee batch instantiation and tokenization results.

Entity	CoffeeBatch_2024_001	Token_Coffee_001
Type	:CollectiveAsset	:Token
Key Properties and Values	:uniqueId "COFFEE-2024-001" :creationDate "2024-06-15T08:00:00Z"	:creationDate "2024-06-16T09:00:00Z" :blockchainTransactionId "tx_coffee_001_abc123"
Governance Linkage	ownedBy Farm_SantaMaria	:represents CoffeeBatch_2024_001 :registeredOnBlockchain Provider_Hyperledger :governedBy DataContract_CoffeeExport

4.2.2. Automated Certification with Governance via Data Contracts

Building on the tokenized asset, we instantiated a verifiable :Certificate (Certificate_Organic_001) as a W3C Verifiable Credential, issued by a trusted certifier and cryptographically linked to the coffee batch. To govern data sharing, a :DataContract (DataContract_CoffeeExport) was created. This contract formally authorized a Consumer_EUImporter to access the batch's data for "EU import compliance," with its terms (purpose, validity, and assignee) encoded in a machine-executable ODRL policy (Table 3). The pipeline's SHACL validation confirmed the structural integrity of both the certificate and contract.

Table 3. Automated governance for coffee export.

Component	AgriTrust Class	Instantiated Example and Purpose	Validation Outcome
Rule Set	:DataContract	*DataContract_CoffeeExport*: Defines the terms for data sharing with the EU importer.	Listed in contract report; SHACL-validated.
Rule Target	:Asset and :Observation	*Covers*: CoffeeBatch_2024_001, Observation_SoilPH. Links governed data to the contract.	Confirmed via :coversAsset and :coversObservation properties.
Rule Assignee	:Agent (DataConsumer)	*Assignee*: Consumer_EUImporter. Identifies who the permissions are granted to.	Correctly identified as the :hasPolicyAssignee.
Executable Logic	:ODRLPolicy	*Linked Policy*: ODRLPolicy_CoffeeExport. Contains machine-readable permissions (e.g., odrl:permission odrl:read).	Presence confirmed via :hasODRLPolicy link.

4.2.3. EUDR Deforestation-Free Verification

EUDR compliance requires a verifiable evidence trail linking the commodity to a geolocated, deforestation-free farm. AgriTrust constructs this by semantically connecting key entities:

- Geolocation: farm_SantaMaria is linked via :hasOperatingLocation to a :Location with a GeoSPARQL-defined coordinate;
- Sustainability credential: the farm holds a Certificate_Sustainability_001 (e.g., "Rainforest Alliance Certified");
- Provenance chain: the coffee batch is :ownedBy the farm, creating an unbroken digital thread.

A compliance officer can execute a single SPARQL query to retrieve this complete dossier: tracing from the token to its asset and to the owning farm, and collecting the farm's location and all associated certificates (Table 4). This transforms a manual documentation exercise into a streamlined, automated verification process.

Table 4. EUDR evidence trail modeled in the ontology.

EUDR Requirement	AgriTrust Ontology Implementation	Instantiated Evidence in Case Study	Validation Outcome
Commodity Identification	:DataContract	Token Coffee_001 represents CoffeeBatch_2024_001.	Pipeline tokenization summary confirms link.
Geolocation of Plot	:Agent (:hasOperatingLocation) → :Location (geo:asWKT)	Farm_SantaMaria operates at Location_SantaMariaFarm (Well-known Text or WKT Point).	Geo-coordinate is present and queryable.
Deforestation-Free Date	Inferred from :Certificate :issuanceDate and linked asset :creationDate.	Farm's Sustainability Certificate issued 15 April 2024. Coffee batch created 15 June 2024.	Dates are machine-readable; logic can check asset date <= certificate date.
Due Diligence System	:DataContract with :purpose and :hasODRLPolicy.	DataContract_CoffeeExport authorizes the importer to access this evidence for compliance.	Contract is valid, has an assignee, and links to an executable policy.

4.2.4. Demonstration of Blockchain-Agnostic Sovereignty

To validate the framework's resistance to vendor lock-in, we instantiated a second coffee batch (CoffeeBatch_2024_002). Crucially, its token (Token_Coffee_002) was :registeredOnBlockchain on Provider_Polygon, a different provider than the first batch. The pipeline's federated query (token_query) retrieved both tokens into a unified view, proving that a data consumer can access and verify assets across different ledgers using a single, standardized AgriTrust interface. The :ownedBy link to Farm_SantaMaria remained im-

mutable, demonstrating that data sovereignty is preserved regardless of the underlying ledger technology (Table 5).

Table 5. Blockchain-agnostic implementation for coffee batches.

Component	Coffee Batch 2024-001	Coffee Batch 2024-002	Validation Outcome
Physical Asset	CoffeeBatch_2024_001 (CollectiveAsset)	CoffeeBatch_2024_002 (CollectiveAsset)	Consistent asset modeling regardless of platform.
Digital Token	Token_Coffee_001 (Token)	Token_Coffee_002 (Token)	The core tokenization function is provider-independent.
Blockchain Provider	Provider_Hyperledger (Enterprise)	Provider_Polygon (Public)	A single producer can utilize different ledgers.
Management Link	:managedByPlatform :Provider_Hyperledger	:managedByPlatform :Provider_Polygon	The :managedByPlatform property decouples logical assets from physical infrastructure.
Governance Status	Linked to DataContract_CoffeeExport.	No linked data contract in current instantiation	Governance (contracts) can be attached as needed, post-tokenization.
Pipeline Validation	Listed under “Hyperledger Fabric” in blockchain report	Listed under “Polygon” in blockchain report	A single query (token_query) retrieved both tokens into a unified view.

The coffee case study demonstrated AgriTrust’s capability to model verifiable provenance for EUDR compliance, with emphasis on geolocation linking and certificate-based attestation. However, coffee represents a relatively straightforward traceability scenario where batches remain identifiable throughout the chain. To test the framework’s flexibility, we next examine a structurally distinct challenge: soybean supply chains where mass balance practices blend commodities from multiple origins.

4.3. Case Study 2: Soybean Supply Chain with Mass Balance and Carbon Metrics

To evaluate AgriTrust’s capacity to handle complex bulk commodity logistics, we instantiated a soybean supply chain. This sector is characterized by the widespread practice of mass balance, where commodities from multiple origins are blended during storage and transport, presenting a distinct challenge for traditional one-to-one traceability models.

4.3.1. Modeling Mass Balance with Cross-Chain Tokens

Soybean batch SoyBatch_2024_001 was modeled as a :CollectiveAsset. Its digital passport, Token_Soy_001, was instantiated as a :CrossChainToken—a specialized class for assets recognized across platforms. While primarily registered on Hyperledger Fabric, it holds a :hasCrossChainReference to an entity registered on Polygon (Table 6). This models a real-world custody transfer (e.g., from a farm-side platform to a port-side platform) without requiring a shared ledger, enabling interoperable, verifiable audit trails for blended commodities.

The soybean case study confirmed that AgriTrust’s CrossChainToken mechanism can accommodate mass balance scenarios through referential linking across platforms. Yet, both coffee and soy involve collective assets (batches). To assess whether the ontology generalizes to fundamentally different asset types requiring individual tracking, we next apply it to beef cattle production.

Table 6. Modeling multi-modal soybean logistics with cross-chain tokens.

Logistical Phase	Physical Reality	AgriTrust Ontology Instantiation	Validation Outcome
Origin and Initial Tokenization	Batch created at origin farm.	SoyBatch_2024_001 (a :CollectiveAsset) is created and :ownedBy Embrapa.	Asset appears in inventory with the correct owner.
Primary Digital Record	First platform issues a digital record.	Token_Soy_001 (a :CrossChainToken) :represents the batch and is registered on Hyperledger Fabric.	Token summary shows token and its blockchain.
Port-Side Acknowledgment	Batch arrives at a port managed by a different platform.	The token :hasCrossChainReference to CrossChainRef_Soy_001, which is registered on Polygon.	Blockchain distribution report shows the CrossChainReference on Polygon.
Mass Balance (Blending)	Conceptually modeled by having multiple :CollectiveAssets as :hasInput to a :Process whose :hasOutput is a new :CollectiveAsset.	Conceptually modeled by having multiple :CollectiveAssets as :hasInput to a :Process whose :hasOutput is a new :CollectiveAsset.	The Process_Harvest model indicates the ontology can link input/output assets.

4.3.2. Integrating Sustainability Performance Data

We instantiated an :EfficiencyMetric (Metric_CarbonFootprint) with a value of "2.3 tonnes CO_2e (carbon dioxide equivalent) per hectare". This metric was linked via :hasEfficiencyMetric to both the Farm_SantaMaria and the Process_Harvest_2024 (Table 7). This integration demonstrates how trusted, granular performance data becomes part of the traceability graph. Such metrics can be governed by data contracts, allowing downstream entities (e.g., sustainable finance institutions) to access verifiable carbon data under explicit, automated terms, creating the foundation for data-driven carbon markets.

Table 7. Integration of carbon footprint metric into the traceability graph.

Component	AgriTrust Class	Instantiated Example and Purpose	Validation Outcome
Performance Indicator	:EfficiencyMetric	Metric_CarbonFootprint: Encodes the KPI "CO_2 Emissions per hectare = 2.3".	Listed in pipeline statistics (2 efficiency metrics in total).
Attribution to Actor	:Agent (DataProducer)	Linked to: Farm_SantaMaria. Attributes the carbon footprint to the farm's operations.	Confirmed via :hasEfficiencyMetric property.
Attribution to Activity	:Process	Also linked to: Process_Harvest_2024. Allows granular analysis per production phase.	Confirmed via :hasEfficiencyMetric property.
Governed Data Access	:DataContract	Potential Coverage: A data contract :coversObservation or can be extended to cover such metrics.	Pipeline validates that contracts successfully govern other linked entities.

Having validated individual animal tracking, we introduce a cross-cutting scenario that tests the ontology's extensibility for precision compliance: agrochemical application tracking for organic certification. This case study focuses not on a new commodity but on the granular process-level data required for automated rule verification across any certified production system.

4.4. Case Study 3: Agrochemical Application Tracking for Certification

We further tested the framework's extensibility and precision by modeling a critical, high-stakes process: agrochemical application tracking for organic certification. This scenario demands granular, time-bound data and automated rule-checking, pushing the ontology's ability to support stringent compliance verification.

Ontology Extension for Agricultural Precision and Automated Verification

The AgriTrust Ontology is built for extensibility. For instance, one may define a new subclass, :SprayApplicationProcess, which inherits from :Process. This allows for modeling specific attributes (e.g., chemical, rate, and date) critical for organic or sustainable certification schemes, without breaking existing interoperability.

Alternatively, a critical organic certification rule to be verified is: "No prohibited substances applied to a certified asset". Using the semantic graph, this is translated into an automated SPARQL query. A certifier's system can query for any certified :Asset and check for linked :Observation nodes where the :observationValue matches a prohibited substance list (Table 8). The pipeline's ability to execute complex cross-entity queries corroborates the feasibility of transforming document-heavy audits into continuous, automated verification.

Table 8. Automated compliance check for certification.

Compliance Rule	Semantic Model Check	Example SPARQL Query Logic	Validation Outcome
Prohibition of specific agrochemicals.	Check if a certified :Asset has an :Observation where :observationValue matches a prohibited substance.	SELECT ?asset WHERE { ?asset a :Asset ; :hasCertificate ?cert ; :hasObservation ?obs . ?obs :observationValue ?val . FILTER (STRSTARTS(?val, "ProhibitedSubstanceX")) }	Returns assets with violations; empty result confirms compliance.
Mandatory pre-harvest intervals.	Check if an :Observation (spray event) date is too close to the asset's harvest :Process date.	SELECT ?asset WHERE { ?asset :hasObservation ?obs . ?obs :observationDate ?obsDate . ?harvestProc :hasOutput ?asset ; a :HarvestProcess . ?harvestProc :endTime ?harvestDate . FILTER (?harvestDate-?obsDate < xsd:dayTimeDuration("P21D")) }	Flags assets sprayed within 21 days of harvest.
Verification of licensed applicators.	Ensure the :Agent linked to the spray :Process (via :managedByPlatform or a new :performedBy property) holds a valid license credential.	The token :hasCrossChainReference to CrossChainRef_Soy_001, which is registered on Polygon.	Confirms operations were performed by authorized personnel.

4.5. Case Study 4: Beef Cattle Production with Individual Tracking and Analytics

Finally, to assess AgriTrust's flexibility in modeling fundamentally different asset types within a single system, we applied it to beef cattle production. This sector requires a dual model: precise individual animal tracking for lifecycle management alongside efficient batch-level processing for operations like feeding and transport.

4.5.1. Individual vs. Collective Asset Models

We instantiated Cattle_001 as an :IndividualAsset, semantically distinct from the :CollectiveAsset used for crops (Table 9). Its token (Token_Cattle_001) was :registeredOnBlockchain on Provider_Ethereum, showcasing how ledger choice can be tailored to the asset's market (e.g., using a public chain for potential decentralized finance integra-

tion or DeFi). This allows a herd management platform and a food safety platform to interoperate using the same ontology.

Table 9. Contrasting individual and collective asset models in the AgriTrust Ontology.

Aspect	Individual Asset Tracking (Beef Cattle)	Collective Asset/Batch Processing (Coffee, Soy)	Validation Outcome
Asset Class	:IndividualAsset (e.g., Cattle_001).	:CollectiveAsset (e.g., CoffeeBatch_2024_001)	Distinct subclasses of :Asset enable clear semantic differentiation.
Token Representation	Token_Cattle_001 represents one specific animal.	A :Token (e.g., Token_Coffee_001) represents an entire batch as a single unit.	The :represents property works identically, but its scope (1:1 vs. 1:many) is defined by the asset class.
Provenance and Processes	Processes (e.g., vaccination, movement) have the individual animal as :hasInput and :hasOutput.	Processes (e.g., harvest, blending) have the entire batch as input/output.	The same :Process class and properties model both, enabling uniform querying of transformation histories.
Key Use Case	Lifecycle passport, genetic tracing, premium "farm-to-fork" branding, disease control.	Mass balance tracking, bulk commodity certification, logistics management.	The framework generalizes traceability concepts to cover diverse sectorial needs.

4.5.2. Growth Performance Analytics Across Livestock Phases

We linked an :EfficiencyMetric (Metric_DailyWeightGain) to a specific :FinishingPhase process, which itself had Cattle_001 as input/output (Table 10). This creates a precise, queryable record of performance contextualized by production phase. This structure enables advanced use cases: genetic benchmarking (comparing animal performance), precision feeding (optimizing diets per cohort), and accurate carbon footprint allocation for Product Environmental Footprint (PEF) studies.

Table 10. Modeling livestock efficiency across production phases.

Production Phase	AgriTrust Process Model	Linked Efficiency Metrics (Examples)	Business Intelligence Application
Rearing/Weaning	Process_Rearing_Phase	Metric_PreWeaningGrowth, Metric_HealthScore	Evaluate dam productivity and calf viability.
Growing/Finishing	Process_Finishing_Phase	Metric_AverageDailyGain, Metric_FeedConversionRatio	Optimize feed costs and time to market.
Health Management	Process_Vaccination	Observation of vaccine batch and date	Ensure compliance with health protocols and trace medication use.
Transport	Process_Transport_ToFarm	Metric_TransportStressIndex (from IoT sensors)	Monitor animal welfare and meat quality predictors.

4.6. Synthesized Cross-Commodity Results and Performance

The case studies collectively provide empirical validation. The Python pipeline served as the concrete integration test, generating definitive quantitative metrics (Table 11).

Table 11. Synthesized quantitative results from the validation pipeline.

Validation Aspect	Result	Interpretation
Knowledge Graph Scale	627 explicit → 2010 triples after reasoning (+220%).	The ontology enables rich deductive querying, deriving implicit knowledge critical for complex supply chain insights.
Semantic Interoperability	7/7 SPARQL modules executed successfully; unified asset inventory query returned coffee, soy, and cattle assets.	A data consumer can obtain a federated view across different commodities and platforms without prior knowledge of proprietary models.
Multi-Provider Operation	Entities successfully distributed and tracked across 3 blockchain providers (Hyperledger, Ethereum, Polygon).	The blockchain-agnostic design is operation-al, preventing vendor lock-in and allowing ledger choice based on cost, jurisdiction, or features.
Governance Enforcement	4/4 governance artifacts (2 certificates, 2 data contracts) passed SHACL validation.	Business rules (sovereignty, contract validity) can be structured and machine-verified, transforming governance into an automatable system feature.
Framework Capability	All core capabilities (tokenization, sovereign ownership, process provenance, verifiable certification) demonstrated across all case studies.	The framework is generalizable and adapts to diverse asset types (individual/collective) and regulatory needs (EUDR, mass balance, organic certification).

4.6.1. Resolving the AgData Paradox

Collectively, the case study results demonstrate how the AgriTrust framework systematically dismantles the tripartite barriers of the AgData Paradox:

- Trust: Explicit :ownedBy properties, verifiable credentials, and blockchain anchoring provide a technical basis for trust. Data contracts automate and render transparent usage terms;
- Interoperability: The shared ontology provides the missing common vocabulary. The pipeline's unified queries are direct proof that heterogeneous data can be integrated seamlessly;
- Governance: Governance is instantiated within the graph as linkable, executable code (via ODRL policies and SHACL), moving it from an abstract principle to a concrete, auditable system feature.

4.6.2. Preliminary Performance and Enforcement Indicators

While comprehensive scalability testing is beyond the scope of this proof-of-concept study, we include runtime measurements for the reasoning and validation pipeline as a concrete illustration of the validation workflow. On a standard laptop (Intel i7, 16GB RAM), OWL 2 RL reasoning was performed on the 627-triple base, completed in 0.34 s, expanding the graph to 2010 triples. SHACL validation of all governance artifacts (two certificates and two data contracts) was completed in 0.12 s with zero violations.

These figures should be interpreted with caution: they reflect the small, hand-crafted nature of the validation dataset and do not represent expected performance at scale. They serve merely as a technical reference point for understanding the pipeline's operation, not as performance baselines for future work.

To demonstrate enforcement capability concretely, we deliberately constructed a malformed data contract instance (DataContract_Invalid) missing the required :validUntil property and lacking a :hasPolicyAssignee. Listing 2 and Table 12 show the resulting SHACL validation error report, which identifies both violations. This illustrates how the framework can prevent non-conformant data from entering the system, providing machine-enforceable governance at the point of data ingestion.

Table 12. Federated SPARQL query results.

Asset	Token	Blockchain	Certificate	CertType
:CoffeeBatch_2024_001	:Token_Coffee_001	:Provider_Hyperledger	:Certificate_Organic_001	:OrganicCertificate
:CoffeeBatch_2024_001	:Token_Coffee_001	:Provider_Hyperledger	:Certificate_Sustainability_001	:SustainabilityCertificate
:Cattle_001	:Token_Cattle_001	:Provider_Ethereum	:Certificate_Health_001	:HealthCertificate

Listing 2. SHACL validation error report for malformed contracts.

```
Validation Report
Conforms: false
Results:
[
   a sh:ValidationResult ;
   sh:focusNode :DataContract_Invalid ;
   sh:resultPath :validUntil ;
   sh:resultSeverity sh:Violation ;
   sh:sourceConstraintComponent sh:MinCountConstraintComponent ;
   sh:value :DataContract_Invalid ;
   sh:resultMessage "Property :validUntil must have exactly 1 value, found 0" ;
]
[
   a sh:ValidationResult ;
   sh:focusNode :DataContract_Invalid ;
   sh:resultPath :hasPolicyAssignee ;
   sh:resultSeverity sh:Violation ;
   sh:sourceConstraintComponent sh:MinCountConstraintComponent ;
   sh:value :DataContract_Invalid ;
   sh:resultMessage "Property :hasPolicyAssignee must have at least 1 value, found 0" ;
]
```

4.6.3. Complex Federated Query Example

To illustrate the framework's query capability across multiple case studies, Listing 3 presents a SPARQL query that retrieves all certified assets from Farm_SantaMaria across different blockchains—a realistic use case for a financial auditor verifying sustainability-linked loans.

Listing 3. Federated SPARQL query for cross-chain certified assets.

```
PREFIX : <https://w3id.org/agrisemantics/AGRITRUST#>
PREFIX geo: <http://www.opengis.net/doc/IS/geosparql/1.1#>

SELECT ?asset ?token ?blockchain ?certificate ?certType WHERE {
   ?asset :ownedBy :Farm_SantaMaria .
   ?token :represents ?asset .
   ?token :registeredOnBlockchain ?blockchain .
   ?asset :hasCertificate ?certificate .
   ?certificate a ?certType .
   FILTER (?certType != :Certificate) # Exclude the base class
} ORDER BY ?asset
```

This query demonstrates how a consumer can obtain a unified view of certified assets across multiple blockchains (Hyperledger and Ethereum) and asset types (coffee batch and individual cattle) without prior knowledge of proprietary data models, precisely the interoperability AgriTrust aims to enable.

These figures should be interpreted with caution: they reflect the small, hand-crafted nature of the validation dataset and do not represent expected performance at scale. They serve merely as a reference point for the scalability studies planned in future work, where systematic benchmarking with larger datasets can quantify reasoning complexity and identify potential bottlenecks. Initial projections based on ontology complexity suggest that reasoning time scales non-linearly with graph size, necessitating optimization strategies such as selective materialization or graph partitioning for production deployments.

In summary, the implementation and initial validation corroborate AgriTrust as a practically viable and scalable blueprint. It has evident potential to successfully transform agricultural data sharing from a trust-based dilemma into a coordinated, rule-based operation, providing the semantic and governance foundation for a federated, equitable agricultural data economy.

5. Discussion

5.1. Threat Model

The AgriTrust ecosystem is designed to be resilient against a range of threats by integrating technical enforcement with its governance model. Table 13 outlines the primary threat actors, their objectives, and the corresponding mitigation strategies.

Table 13. Threat Model and Mitigation Strategies of the AgriTrust Framework.

Threat Actor	Primary Goal	Mitigation Strategy	Technical Implementation
Malicious Data Consumer	To use data in ways that violate the agreed-upon terms, such as for unauthorized commercial analysis or reselling.	Machine-Readable Data Contracts and Automated Enforcement. Data usage policies are not just legal text but are codified in machine-readable contracts.	Contracts explicitly define :allowedUsage and :prohibitedUsage. Access control lists (ACLs) for APIs and SPARQL endpoints are automatically generated from these terms. Every data query is logged against its corresponding contract ID for full auditability and compliance monitoring.
Dishonest Data Producer	To introduce fraudulent data into the system to gain an unfair advantage, such as claiming false sustainable certification for a product.	Cryptographic Data Provenance and Multi-Source Verification. The system establishes a verifiable chain of custody for all data.	Every data observation is digitally signed with the producer's (or their IoT device's) private key, creating an uncontested proof of origin. For high-stakes claims (e.g., geographic origin), data is automatically cross-referenced with trusted independent sources, such as satellite imagery platforms. The framework enables platforms to implement such verification by providing access to the necessary data via data contracts.
Rogue Platform Provider	To operate a node that deliberately violates governance rules, for example, by serving tampered data or ignoring revocation signals.	On-Chain Integrity Anchors and Governance Authority Oversight. A decentralized ledger provides a neutral ground for verifying critical system metadata.	Essential identifiers and hashes (for tokens, assets, and contracts) are registered on a blockchain, creating an immutable integrity anchor. The governance authority maintains the power to de-list non-compliant platforms, while clients can independently verify any data asset against its on-chain hash.

Table 13. *Cont.*

Threat Actor	Primary Goal	Mitigation Strategy	Technical Implementation
External Attacker	To compromise the classic security triad: confidentiality, integrity, or availability of the system and its data.	Standard Cybersecurity Hardening and Decentralized Architecture. Defense-in-depth principles are applied to the federated structure.	All data-in-transit is protected with modern encryption (HTTPS/TLS 1.3). Platform code undergoes regular security audits. Crucially, the federated design eliminates single points of failure, inherently containing the impact of incidents like a Distributed Denial-of-Service (DDoS) attack on any single node.

5.2. Technical Enablers: Balancing Transparency and Privacy

The mitigations outlined above are enabled by two foundational technical choices: the on-chain/off-chain data partitioning model and the machine-executable governance layer.

On-chain vs. off-chain data model. Only integrity-critical metadata is stored on the blockchain: hashes of data contracts, tokens, and assets, along with transaction IDs and timestamps. This data is immutable and publicly verifiable but reveals no sensitive business information. All detailed, commercially sensitive data, including financial terms, operational yields, personal identifiers, and raw sensor data, resides off-chain in the respective TaaS platforms' secured databases. Access to this off-chain data is strictly governed by data contracts, ensuring that only authorized parties can view specific data points for specific purposes and durations.

Machine-executable governance. Data contracts are not legal text but are codified in machine-readable ODRL policies linked to on-chain integrity anchors. This enables automated enforcement: API gateways and SPARQL endpoints derive access control lists directly from contract terms, and every query is logged against its contract ID for full auditability. Cryptographic signatures on data observations provide uncontested proof of origin, while multi-source verification (e.g., cross-referencing geolocation claims with satellite imagery) adds an additional layer of trust for high-stakes assertions.

5.3. Trust Architecture: Distributed Anchors and Governance

The AgriTrust framework does not eliminate trust but rather distributes and formalizes it. It replaces the need to trust any single participant with verifiable trust in three interdependent anchors:

1. Cryptographic: The immutability of underlying blockchains provides a verifiable "root of trust" for the existence and state of tokens, contracts, and assets. Clients can independently verify any digital asset against its on-chain hash, ensuring data integrity without relying on platform providers.
2. Semantic: The shared, governance-maintained core ontology ensures that all participants have a common understanding of what the data means. This prevents semantic attacks and misinterpretation, creating a trusted foundation for interoperability across platforms and jurisdictions.
3. Institutional: The consortium-based governance authority acts as an ultimate arbiter, capable of auditing platforms, revoking accreditations, and resolving disputes. While the technical anchors provide verifiability, the institutional anchor provides accountability and recourse, enforcing the rules of the ecosystem through legitimate authority.

It is important to distinguish the trust the framework creates from the trust it requires. AgriTrust provides cryptographic verifiability for all data once it enters the system, creating an immutable chain of provenance and consent. However, it relies on established procedures and trusted third parties (e.g., certifiers and satellite imagery providers) to bootstrap

the initial trust in raw data at the point of origin. The framework's role is to make this initial attestation and all subsequent transformations transparent and auditable.

Together, these three anchors create a layered trust model where technical verifiability, semantic alignment, and institutional oversight reinforce one another. This makes it economically and technically prohibitive to attack the system at scale, while providing clear recourse and audit trails for any malicious activity.

5.4. Economic, Environmental, and Social Benefits

The automated, pre-authorized data sharing for certifications (e.g., sustainable and EUDR) drastically reduces the need for costly and disruptive manual audits. Certifiers can verify claims remotely and at scale. Alternatively, the framework provides the technological substrate for producers to reliably prove compliance with stringent market standards (e.g., EUDR and carbon footprint), unlocking access to premium markets and price incentives. Moreover, the "equitable value sharing" pillar, operationalized through data contracts, makes a formal mechanism for producers to be compensated for the value their data creates, for example, through automatic premium payments upon successful certification.

In terms of environmental compliance, integrating geospatial data (e.g., farm boundaries) with immutable tokenized assets provides cryptographically verifiable proof of deforestation-free supply chains, directly addressing regulatory requirements like the EUDR. In parallel, the ability to track inputs and outputs across processes enables the calculation of efficiency metrics like carbon footprint and water usage. This data empowers producers to identify inefficiencies and optimize resource use (management), reducing the environmental impact of production.

The principle of data sovereignty shifts power dynamics, ensuring that producers are active participants and beneficiaries in the digital ecosystem. In addition, consumers and retailers gain unprecedented visibility into the origin and production practices of their food, fostering trust and enabling informed purchasing decisions.

To contextualize these potential impacts and AgriTrust's unique value proposition, it is instructive to position its integrated approach against the landscape of existing, more isolated technological solutions. The following analysis highlights how AgriTrust synthesizes its strengths while systematically addressing its characteristic weaknesses.

5.5. Comparison with Existing Approaches

AgriTrust stands apart by integrating formal mechanisms into a coherent, principled architecture. The following discussion positions the framework against the major categories of related work, highlighting how it synthesizes their strengths and mitigates weaknesses (Table 14).

Table 14 reframes the comparison around the question: "What specific barriers of the AgData Paradox does each approach address, and what critical barriers remain?" Isolated blockchain platforms address trust through immutability but leave interoperability and governance barriers intact because they secure data within silos rather than enabling cross-silo sharing. Centralized platforms offer operational efficiency and governance within their boundaries but inherently concentrate control, creating new trust deficits around data sovereignty and fair value sharing. Semantic ontologies provide the vocabulary for interoperability but stop at description, lacking the machinery for governance execution. Data space architectures such as IDS and CEADS articulate the correct principles of sovereignty, federation, and policy-based access but remain at the reference architecture level, requiring sector-specific instantiation with concrete semantics.

Table 14. Comparison of AgriTrust with existing approaches.

Approach Category	Isolated Blockchain Platforms	Centralized Agri-Platforms	Semantic Ontologies	Data Space Architectures
Specific System Example	IBM Food Trust, BeefChain, TE-FOOD	AgroTools, JohnDeere API, proprietary cooperative platforms	AGROVOC, FOODON, Plant Ontology, Soares (2025) framework	International Data Spaces (IDS), CEADS, Pontus-X
Core Strength	Strong data integrity and immutability within a single supply chain	Operational efficiency, data aggregation within walled garden	Shared vocabulary, semantic harmonization	Sovereign federated sharing blueprint, decentralized identity
AgData Paradox Barriers Addressed	Data integrity	Usability	Interoperability (syntactic/semantic)	Sovereignty principles Federation model
Critical Barriers Remaining	Cross-platform interoperability Formal governance model for data sovereignty Value sharing mechanisms	Vendor lock-in, data portability Producer sovereignty technically unenforced Benefit sharing opaque	Governance artifacts not modeled No executable policies Tokenization semantics absent	Sector-specific implementation missing Concrete semantics for tokenization and contracts Machine-executable governance
How AgriTrust Extends	Adopts blockchain as integrity anchor but embeds assets in blockchain-agnostic semantic layer; adds explicit governance classes (DataContract, Certificate)	Architecturally enforces producer sovereignty via ownedBy property; shared ontology enables data portability across competing platforms	Extends foundational ontologies (PROV-O, SOSA/SSN) with explicit governance artifacts (DataContract, Certificate) linked to W3C VC and ODRL standards, bridging description to automated execution	Provides concrete agricultural instantiation of data space principles; adds domain ontology with tokenization, traceability, and governance semantics aligned with IDS reference architecture

AgriTrust engages critically with the data space literature by asking: What does it take to move from the IDS reference architecture to a functioning agricultural data space? The answer, as our framework demonstrates, is a domain ontology that explicitly models the artifacts needed for operational governance—tokens that represent assets, contracts that encode usage policies, and certificates that provide verifiable attestations—and links them to executable standards (ODRL and W3C VC). The specific challenges encountered in this translation include (1) balancing expressivity with computational tractability in OWL, (2) designing for multi-provider extensibility without fragmentation, and (3) embedding sovereignty as a first-class property (ownedBy) that cannot be overridden by platform logic. These challenges, and AgriTrust's design responses to them, illustrate what a concrete, sectoral implementation of data space principles entails.

5.6. Preliminary Evaluation and Path to Future Validation

Case studies with multiple commodities serve as a functional proof of concept (PoC), validating the core architecture of the AgriTrust framework and its semantic interoperability capabilities. This PoC-based evaluation specifically tested theorem-type environments (including propositions, lemmas, corollaries, etc.), which can be formatted as follows:

- Semantic correctness and reasoning: the ability of the AgriTrust Ontology to formally model governance assets, processes, and artifacts and to support OWL 2 RL reasoning (expanding the graph from 627 to 2010 triples);
- Federated query capability: the feasibility of executing unified SPARQL queries in a simulated cross-platform and cross-blockchain environment to retrieve a consolidated view of assets, provenance, and contracts;
- Governance artifact instantiation: the technically correct instantiation and linking of key governance classes (:DataContract and :Certificate) with executable policies (ODRLs) and verifiable credentials;
- The agnostic reference to blockchain: the ability of the ontological model to represent and link assets registered across different simulated blockchain providers (Hyperledger Fabric, Ethereum, and Polygon).

However, this proof of concept (PoC) is inherently limited in scope and scale. It operates on a static, synthesized dataset and does not assess performance under production conditions. Therefore, a full quantitative assessment of the framework's operational performance [31], economic impact, and adoption dynamics remains crucial for future work. The current assessment is limited by its scale (using limited simulated data), its scope (lack of real-time performance testing under high load and complex interactions between multiple stakeholders), and the preliminary nature of stakeholder feedback.

Beyond technical and legal challenges, the framework's adoption faces significant socio-technical hurdles that must be addressed for equitable inclusion. First, incentive alignment for smallholders: Producers with limited digital literacy or connectivity may perceive data sharing as risky with unclear benefits. The framework must demonstrate tangible value propositions, such as reduced audit costs, premium market access, or direct compensation via data contracts, that outweigh perceived risks. Future work should explore low-tech interfaces (USSD and SMS) and asynchronous protocols that function in low-connectivity environments, as well as cooperative-level onboarding models where producer associations intermediate digital services.

Furthermore, the consortium-based governance authority described in Section 3.2.1 requires legitimate representation from diverse stakeholders: producer associations, industry, certifiers, government, and civil society. Realistically constituting and funding such an entity poses political and economic challenges. Who bears the initial setup costs? How are voting rights allocated? What mechanisms prevent regulatory capture by dominant actors? While these questions extend beyond technical design, they are critical to the framework's viability. Pilot implementations with existing cooperative structures or multi-stakeholder initiatives (e.g., roundtables for sustainable commodities) could provide experimental grounds for testing governance models.

In addition, the persistent connectivity gap in rural Brazil means that universal participation cannot be assumed. Beyond technical solutions, this requires institutional innovation: partnerships with extension services, integration with existing paper-based traceability systems during transition periods, and capacity-building programs. The framework's design, with sovereignty preserved locally, supports gradual, opt-in adoption, but deliberate socio-technical strategies are needed to ensure that the data economy does not exclude the very producers whose participation is essential for supply chain transparency.

Validation Scale and Scope

A primary limitation of the current study is the scale and controlled nature of the validation dataset. With only 627 explicit triples and 4 physical assets, the knowledge graph cannot simulate the complexity, volume, or inconsistency of real-world agricultural data. This limitation has two important consequences.

First, scalability remains untested. While the ontology's logical structure suggests it can accommodate large-scale deployments—classes are designed for extensibility, and properties support graph traversal—empirical validation with datasets containing thousands or millions of assets is necessary to quantify reasoning performance, query latency, and system throughput under load.

Second, robustness to real-world inconsistencies cannot be assessed. The hand-crafted dataset was designed to conform perfectly to the ontology, by construction revealing no modeling blind spots. Real-world data, however, arrives with errors, omissions, and semantic ambiguities that may expose limitations in the ontology's axiomatization or the SHACL constraints' coverage.

Recognizing these limitations, we have initiated a program of work to systematically evaluate AgriTrust's scalability and robustness. Forthcoming research requires

1. The generation of synthetic datasets (e.g., ranging from 10 to 1,000,000 assets), spanning multiple supply chains and governance scenarios;
2. The measurement of OWL reasoning time, SHACL validation throughput, and SPARQL query latency as functions of dataset size;
3. An analysis of memory utilization and the identification of performance bottlenecks;
4. Testing of the framework's behavior in the presence of intentionally introduced data inconsistencies (missing required properties, invalid references, and malformed policies).

Additionally, ongoing partnerships with Brazilian cooperatives and agribusinesses can enable pilot deployments with real operational data, providing the ultimate test of the framework's robustness to real-world conditions. These deployments can also generate empirical evidence on adoption dynamics, user acceptance, and the practical value proposition for producers, dimensions that cannot be captured through synthetic validation alone.

5.7. Additional Limitations

The framework's adoption faces significant real-world constraints. First, it presupposes a baseline of digital literacy and connectivity among producers. The persistent digital divide in rural areas, particularly in developing regions like Brazil, remains a formidable barrier to universal inclusion. To ensure equitable access, future work must investigate low-tech interfaces (e.g., Unstructured Supplementary Service Data (USSD) and Short Message Service (SMS)) and asynchronous, offline-capable protocols that can bridge connectivity gaps, as suggested in Section 4.4.

Furthermore, establishing the operational ecosystem requires substantial upfront investment and coordination. The formation of a multi-stakeholder governance authority, the development of core platform infrastructure, and the initial onboarding of participants all incur significant costs. A clear, phased rollout strategy is therefore essential to demonstrate incremental value and build momentum.

Finally, the framework must be legally and semantically adaptable. It will need to evolve in tandem with changing data privacy laws and agricultural regulations. This necessitates ongoing engagement with policymakers and a core ontology designed for extensibility, ensuring the system remains compliant and relevant over time.

6. Conclusions

This study addresses a fundamental limitation of current agricultural information systems: the inability to support interoperable, verifiable, and machine-actionable data exchange across heterogeneous production, certification, and market environments. This limitation is conceptualized here as the "AgData Paradox", in which the immense potential of agricultural data remains locked in fragmented silos due to mistrust and a lack of interoperability. Our research demonstrates a coherent architectural blueprint: by deeply

integrating a principled, multi-stakeholder governance model with a formal semantic layer, realized through the AgriTrust Ontology within a blockchain-agnostic, federated architecture, it is possible to specify how agricultural data sharing could be transformed from a trust-based dilemma into a governed, automated operation. The AgriTrust framework specifies that technical execution should inherently enforce the fair business rules of data sovereignty, transparent contracts, equitable value sharing, and regulatory compliance. The semantic reasoning validation across coffee, soybean, and beef supply chains, processing a proof-of-concept knowledge graph of 2010 triples, confirms that the ontology can represent assets across three distinct simulated blockchain providers, anticipating initial evidence of the framework's conceptual viability. However, as discussed in Section 5.6, this represents a proof of concept; real-world deployment with production-scale data and live blockchain integration remains essential to validate the framework's operational claims.

The quantitative validation across the simulations of coffee, soybean, and beef supply chains, processing a federated knowledge graph of 2010 triples and managing assets across three distinct blockchain providers, gives concrete evidence of the framework's viability. This carries significant, practical implications for all stakeholders. For data producers, including farmers and cooperatives, AgriTrust technically enforces sovereignty, transforming them from passive data sources into empowered economic agents. It drastically reduces the cost and burden of compliance audits through automation and creates the infrastructure for novel revenue streams from governed data sharing. For technology providers and certifiers, the shared semantic layer and standardized interfaces drastically lower the cost and complexity of ecosystem integration. It enables certifiers to perform scalable, remote, and automated compliance verification, enhancing market integrity while reducing their operational overhead. For global markets and consumers, the framework delivers the foundational trust required for a modern, transparent food system. It provides cryptographically verifiable proof of origin, sustainability, and ethical practices, enabling informed consumer choices and supporting the premium value chains required by regulations like the EUDR.

Translating this proven architectural blueprint into widespread industry impact requires focused future work. We outline a concrete pathway organized around scaling the infrastructure, deepening its capabilities, and broadening its accessibility. The immediate priority is to move from proof of concept to demonstrating scalability and economic robustness through phased implementations with partner cooperatives and agribusinesses to operate AgriTrust in real-world environments, handling real volumes of transactions and user interactions. Furthermore, it will be necessary to provide systematic measurements of federated query latency, blockchain transaction throughput, and system resilience under load, based on the initial metrics of this study. Additionally, a longitudinal study must be conducted to quantify the framework's effect on key metrics: reduced audit costs; increased revenue from new data for producers; reduced integration costs; and time to market for service providers.

To unlock more advanced functionality, future research should integrate complementary technologies. Privacy-preserving analytics, by developing interfaces for secure multi-party computation or federated learning, can enable collaborative insights and benchmarking across the ecosystem without compromising individual data sovereignty. Data-driven financial instruments, integrating the framework's verifiable data and settlement layer with DeFi primitives, can pioneer new instruments such as parametric crop insurance, carbon credit futures, or supply chain financing based on real-time, trusted performance data. Advanced cross-chain orchestration, investigating specialized protocols such as cross-chain relays and atomic swaps, can enable complex, atomic transactions across different ledgers, moving beyond referential interoperability to full logical interoperability for business processes.

To achieve inclusive adoption, the framework must evolve to meet diverse real-world constraints. Ontology expansion should extend the core ontology to cover more agricultural sectors, including horticulture, dairy, and aquaculture, and additional value-chain processes. A developer ecosystem with standardized APIs must provide robust software development kits and user-friendly interfaces to lower the barrier to entry for new platform and application developers. Finally, bridging the digital divide requires future work to design low-tech interfaces, such as USSD, SMS, and offline-capable protocols, along with novel onboarding models to ensure producers with limited connectivity or digital literacy can participate in and benefit from the data economy.

The AgriTrust framework provides a viable and coherent pathway toward a transparent, efficient, and equitable agricultural data economy. By addressing the core technical and economic barriers of the AgData Paradox through integrated governance and semantics, it establishes the necessary foundation to ensure that the value generated by the sector's digital transformation is fairly shared among all participants. This work unlocks data's potential to drive the sustainability, resilience, and fairness required for the future of global food systems. The next steps, centered on scaling, deepening, and broadening, chart a clear research and development agenda to translate this blueprint into tangible impact. Foremost among these is systematic scalability validation with larger datasets (work already underway in complementary studies) and real-world deployments to move from proof of concept to operationally validated systems.

Author Contributions: Conceptualization, I.B.; methodology, I.B. and F.M.S.; software, I.B. and F.M.S.; validation, I.B., D.D. and F.M.S.; formal analysis, I.B.; investigation, I.B., J.G.A.B. and É.L.B.; resources, I.B.; data curation, I.B. and F.M.S.; writing—original draft preparation, I.B.; writing—review and editing, J.G.A.B., É.L.B., D.D. and F.M.S.; visualization, I.B.; supervision, J.G.A.B.; project administration, J.G.A.B.; funding acquisition, J.G.A.B. and É.L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research and ACP were funded by FAPESP, grant number 2022/09319-9.

Data Availability Statement: This article is an improved version of the work initially presented by the first author in the preprint arXiv:2511.05572. The updated ontology is available through AgroPortal (https://w3id.org/agrisemantics/AGRITRUST, accessed on 13 March 2026), which provides content negotiation and serves multiple formats (json, rdf/xml, text/turtle, text/n3, and text/html). The readers may access the simulated data and Python pipeline files used for ontology validation here: https://www.gitlab.cnptia.embrapa.br/m327882/agritrust-ontology/-/tree/main/Paper, accessed on 13 March 2026.

Acknowledgments: This article is an expanded version of the work initially presented in the preprint arXiv:2511.05572. I.B. is thankful for the fundamental assistance of generative AI (DeepSeek-V3-0324 and ChatGPT-4o at earlier stages) in achieving new knowledge via human–machine controlled interactions. All AI outputs were reviewed, validated, and refined by the authors, who assume full responsibility for the final framework. É.L.B. thanks the National Council for Scientific and Technological Development (CNPq)/Research Productivity Fellowship (PQ). The authors also thank the relevant contributions provided by Clement Jonquet (https://agroportal.eu/, accessed on 13 March 2026) and four independent reviewers, as well as the fruitful interactions with colleagues from the Center for Science in the Development of Digital Agriculture (https://www.semear-digital.cnptia.embrapa.br/en/, accessed on 13 March 2026) and GO FAIR Agro Brasil Network (https://go-fair-agro.github.io/, accessed on 13 March 2026) for the enlightening discussions, knowledge exchange, and collaboration opportunities.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Massruhá, S.M.F.S.; Leite, M.D.A.; Luchiari Junior, A.; Evangelista, S.R.M. Digital transformation in the field towards sustainable and smart agriculture. In *Digital Agriculture: Research, Development and Innovation in Production Chains*; Massruhá, S.M.F.S., Leite, M.A.A., Oliveira, S.R.M., Meira, C.A.A., Luchiari Junior, A., Bolfe, É.L., Eds.; Embrapa: Brasília, Brazil, 2023; Chapter 1; pp. 17–35. Available online: https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1156698/1/LV-Digital-agriculture-2023-cap1.pdf (accessed on 13 February 2026).
2. Bolfe, E.L.; Barbedo, J.G.A.; Massruhá, S.M.F.S.; de Souza, K.X.S.; Assad, E.D. Challenges, trends and opportunities in digital agriculture in Brazil. In *Digital Agriculture: Research, Development and Innovation in Production Chains*; Massruhá, S.M.F.S., Leite, M.A.A., Oliveira, S.R.M., Meira, C.A.A., Luchiari Junior, A., Bolfe, É.L., Eds.; Embrapa: Brasília, Brazil, 2023; Chapter 16; pp. 281–299. Available online: https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1156772/1/LV-Digital-agriculture-2023-cap16.pdf (accessed on 13 February 2026).
3. Oliveira, S.E.M.; Nakagawa, L.; Lopes, G.R.; Visentin, J.C.; Couto, M.; Silva, D.E.; d'Albertas, F.; Pavani, B.F.; Loyola, R.; West, C. The European Union and United Kingdom's deforestation-free supply chains regulations: Implications for Brazil. *Ecol. Econ.* **2024**, *217*, 108053. [CrossRef]
4. Wiseman, L.; Sanderson, J.; Zhang, A.; Jakku, E. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS-Wagening. J. Life Sci.* **2019**, *90–91*, 100301. [CrossRef]
5. Archer, M.; Ravn, L.; Thylstrup, N.B. The political economy of platformed silos: Theorizing data storage reconfigurations in the age of interoperability capitalism. *Big Data Soc.* **2025**, *12*, 20539517241303144. [CrossRef]
6. Bergier, I.; Bolfe, É.L.; Drucker, D.P.; Inamasu, R.Y.; Santos, P.M.; De Abreu, U.G.P.; Da Silva, T.L.; Carvalho, F.D.F.; Romani, L.A.S.; Barbedo, J.G.A.; et al. Data reporting in agri-food platforms: Sharing, privacy, consumer demand, and the role of public policies. *Pesqui. Agropecu. Bras.* **2025**, *60*, e04113. [CrossRef]
7. Bergier, I. AgriTrust: A Federated Semantic Governance Framework for Trusted Agricultural Data Sharing. *arXiv* **2025**, arXiv:2511.05572. [CrossRef]
8. Kamilaris, A.; Fonts, A.; Prenafeta-Boldú, F.X. The rise of blockchain technology in agriculture and food supply chains. *Trends Food Sci. Technol.* **2019**, *91*, 640–652. [CrossRef]
9. Majumdar, P.; Mitra, S. Blockchain technology for society 4.0: A comprehensive review of key applications, requirement analysis, research trends, challenges and future avenues. *Clust. Comput.* **2024**, *27*, 7059–7081. [CrossRef]
10. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [CrossRef]
11. Antoniou, G.; Van Harmelen, F. Web Ontology Language: OWL. In *Handbook on Ontologies, International Handbooks on Information Systems*; Staab, S., Studer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2009. [CrossRef]
12. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
13. Carroll, S.R.; Garba, I.; Figueroa-Rodríguez, O.L.; Holbrook, J.; Lovett, R.; Materechera, S.; Parsons, M.; Raseroka, K.; Rodriguez-Lonebear, D.; Rowe, R.; et al. The CARE Principles for Indigenous Data Governance. *Data Sci. J.* **2020**, *19*, 43. [CrossRef]
14. Global Data Alliance. Trust Across Borders, Cross-Border Data Policy Index. Available online: https://globaldataalliance.org/wp-content/uploads/2023/07/07192023gdaindex.pdf (accessed on 13 March 2026).
15. Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge Graphs. *ACM Comput. Surv.* **2021**, *54*, 1–37. [CrossRef]
16. Knublauch, H.; Kontokostas, D. Shapes Constraint Language (SHACL). W3C Recommendation. 2017. Available online: https://www.w3.org/TR/shacl/ (accessed on 13 March 2026).
17. Sporny, M.; Noble, G.; Longley, D.; Burnet, D.C.; Zundel, B.; Hartog, K.D. Verifiable Credentials Data Model 1.1. W3C Recommendation. 2022. Available online: https://www.w3.org/TR/vc-data-model-1.1/ (accessed on 13 March 2026).
18. Iannella, R.; Villata, S. ODRL Information Model 2.2. W3C Recommendation. 2018. Available online: https://www.w3.org/TR/odrl-model/ (accessed on 13 March 2026).
19. Albertoni, R.; Brownning, D.; Cox, S.J.D.; Bentran, A.G.; Perego, A.; Winstanley, P. Data Catalog Vocabulary (DCAT)—Version 3. W3C Recommendation. 2024. Available online: https://www.w3.org/TR/vocab-dcat-3/ (accessed on 13 March 2026).
20. Jakku, E.; Taylor, B.; Fleming, A.; Mason, C.; Fielke, S.; Sounness, C.; Thorburn, P. "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in Smart Farming. *NJAS-Wagening. J. Life Sci.* **2019**, *90–91*, 100285. [CrossRef]
21. Bergier, I.; Barbedo, J.G.A.; Bolfe, É.L.; Romani, L.A.S.; Inamasu, R.Y.; Massruhá, S.M.F.S. Framing Concepts of Agriculture 5.0 via Bipartite Analysis. *Sustainability* **2024**, *16*, 10851. [CrossRef]
22. Jonquet, C.; Toulet, A.; Arnaud, E.; Aubin, S.; Yeumo, E.D.; Emonet, V.; Graybeal, J.; Laporte, M.A.; Musen, M.A.; Pesce, V.; et al. AgroPortal: A vocabulary and ontology repository for agronomy. *Comput. Electron. Agric.* **2018**, *144*, 126–143. [CrossRef]
23. Grati, R.; Fattouch, N.; Boukadi, K. Ontologies for Smart Agriculture: A Path Toward Explainable AI—A Systematic Literature Review. *IEEE Access* **2025**, *13*, 72883–72905. [CrossRef]

24. Soares, F.M.A. Semantic Interoperability Framework for Data-Centric Applications in Agriculture. Ph.D. Thesis, University of São Paulo, São Paulo, Brazil, University of Twente, Enschede, The Netherlands, 2025; 280p. [CrossRef]
25. *ISO/TC 347*; Standardization in the Area of Data-Driven Agrifood Systems. International Organization for Standardization: Geneva, Switzerland, 2023. Available online: https://www.iso.org/committee/9983782.html (accessed on 13 March 2026).
26. Tian, F. An agri-food supply chain traceability system for China based on RFID & blockchain technology. In Proceedings of the 2016 13th International Conference on Service Systems and Service Management (ICSSSM), Kunming, China, 24–26 June 2016; pp. 1–6. [CrossRef]
27. Pookkaman, W.; Samanchuen, T. An Innovation Framework of Medical Organic Cannabis Traceability in Digital Supply Chain. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 196. [CrossRef]
28. Otto, B.; Hompel, M.T.; Wrobel, S. International Data Spaces. In *Digital Transformation*; Neugebauer, R., Ed.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 109–128. [CrossRef]
29. CEADS—Common European Agricultural Data Space. Trust in Data. Growth in Agriculture. 2026. Available online: https://ceads.eu (accessed on 19 January 2026).
30. CEADS—Common European Agricultural Data Space. Data Spaces Support Centre. Data Spaces Blueprint v2.0. 2026. Available online: https://dssc.eu/space/BVE2/1071251457/Data+Spaces+Blueprint+v2.0+-+Home (accessed on 19 January 2026).
31. Westerkamp, M.; Küpper, A. Instant Function Calls Using Synchronized Cross-Blockchain Smart Contracts. *IEEE Trans. Netw. Serv. Manag.* **2023**, *20*, 2136–2150. [CrossRef]