# Bringing Semantics into the Common European Agricultural Data Space: (Meta)data enrichment with controlled vocabularies in Pontus-X

Filipi Miranda Soares[1,*], Roberto García[2], Miguel Pincheira[3], Raffaele Giaffreda[3] and Clement Jonquet[1,*]

[1]*MISTEA, University of Montpellier, INRAE & Institut Agro, 2 place Pierre Viala, 34060 Montpellier, France*
[2]*Universitat de Lleida, Computer Engineering and Digital Design Department, Jaume II 69, 25001, Lleida, Spain*
[3]*Fondazione Bruno Kessler, OpenIoT Research Unit, Via Sommarive 18, 38123 Trento, Italy*

**Abstract**

The Common European Agricultural Data Space (CEADS) aims to enable trusted, interoperable data sharing across the agri-food sector through a federated ecosystem of tools, services and Data Sharing Initiatives (DSIs). However, achieving "meaningful", i.e., semantic, interoperability requires not only governance and technical alignment but also shared, machine-interpretable semantics that support discovery, integration, and reuse of heterogeneous data assets (datasets, algorithms, and services). In CEADS, this is being realised through the adoption of AGROVOC as the standard vocabulary for the thematic description of datasets, together with the GeoDCAT Application Profile for geospatial metadata. This paper presents three preliminary complementary contributions implemented across two Pontus-X-based DSIs participating in CEADS: AgrospAI and AgriFoodTEF. First, AgroAnnotator performs ontology-based annotation of agricultural documents and datasets in AgrospAI, relying on AgroPortal Annotator within a trusted compute-to-data environment to produce reusable semantic annotations. Second, two publication-time metadata enrichment prototypes support controlled-term selection (i.e., semantic tagging) for data asset descriptions from AGROVOC or any other controlled-vocabularies served through AgroPortal vocabulary hub. Third, Pontus-X-native metadata are exposed through a GeoDCAT-AP wrapper in AgrospAI and a DCAT-compatible catalogue endpoint in AgriFoodTEF. Technical validation is provided through preliminary experiments that demonstrate the publication and cataloguing workflows. Collectively, these early contributions demonstrate practical pathways for embedding vocabulary-driven semantics within federated data spaces, providing a foundation for CEADS semantic interoperability and potentially reusable by other data spaces.

**Keywords**

Semantic Interoperability, Data Spaces, Ontologies, Controlled Vocabularies, AI Ready Data, Dataset Application Profiles, Pontus-X, AgroPortal, DCAT-AP, GeoDCAT-AP

## 1. Introduction

The Common European Agricultural Data Space (CEADS)[1] is being developed to enable trusted data sharing and reuse across the European agri-food sector through a federated ecosystem of platforms and services as part of the European strategy for sectoral data spaces [1]. Within this emerging architecture, participating regional or project specific "data spaces" such as AgrospAI[2] AgrifoodTEF Data Space[3] are commonly referred to as Data Sharing Initiatives (DSIs) to distinguish them from the overarching common, federated data space, i.e., CEADS itself. Although CEADS is still under construction and its

*Manuscript submitted to the NeXt-generation Data Governance workshop (NXDG 2026), co-located with 22nd SEMANTiCS, 15-17 September 2026 Ghent, Belgium*
*Corresponding author.
filipi.miranda-soares@inrae.fr (F. M. Soares); roberto.garcia@udl.cat (R. García); mpincheiracaro@fbk.eu (M. Pincheira); rgiaffreda@fbk.eu (R. Giaffreda); clement.jonquet@inrae.fr (C. Jonquet)
https://cv.hal.science/fmirandasoares (F. M. Soares); https://rhizomik.net/~roberto/ (R. García); https://jonquet.mystrikingly.com/ (C. Jonquet)
0000-0002-0674-7960 (F. M. Soares); 0000-0003-2207-9605 (R. García); 0000-0002-5847-2145 (M. Pincheira); 0000-0002-3075-0932 (R. Giaffreda); 0000-0002-2404-1582 (C. Jonquet)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]https://ceads.eu
[2]https://portal.agrospai.udl.cat/
[3]https://dataspace.agrifoodtef.eu/

federated catalogue and interoperability services are not yet operational, its design is guided by the recommendations of the International Data Spaces Protocol [2] and the Data Space Support Centre (DSSC) Blueprint [3]. These recommendations emphasize the importance of shared data models and semantic interoperability mechanisms, including the adoption of standardized vocabularies such as the Data Catalog Vocabulary (DCAT)[4] for catalogue interoperability, as well as the use of vocabulary hubs or services (Section 1.5 of [3]) to support consistent, machine-interpretable descriptions across DSIs. Furthermore, services and tools that support these capabilities, such as AgroPortal/OntoPortal, described later in this paper, are often registered in the DSSC Blueprint Data Spaces Toolbox.[5]

Despite these recommendations in governance and technical architecture, semantic interoperability remains a central challenge for CEADS and, more generally, for the other Common European Data Spaces [4]. Data assets (datasets, algorithms, and services) may become technically accessible across DSIs, yet remain difficult to discover, compare, and reuse if their metadata and content descriptions lack shared, machine-interpretable meaning. Two complementary semantic challenges are particularly visible in federated data spaces.

First, many datasets are published with limited or inconsistent descriptive information, making data discovery [4] reliant on literal keyword matching rather than concept-based search grounded in domain ontologies or controlled vocabularies, an issue also highlighted in related studies [5, 6]. Second, even when platforms maintain rich internal metadata representations, schema alignment and semantic transformation [4] remain challenging; differences in DSIs' catalogue models and profiles hinder cross-DSI interoperability and limit effective data exchange. Addressing these barriers requires practical mechanisms both to enrich dataset content with semantic descriptors and to expose platform-specific metadata through standards-compliant representations, challenges likewise discussed in prior work [6, 7, 8].

This paper presents prototypes for bringing semantics into CEADS through complementary implementations deployed in AgrospAI and AgriFoodTEF, two Pontus-X-based DSIs participating in CEADS. The proposed components address two layers of semantic interoperability: enrichment of dataset content through ontology-based annotation, and harmonization of catalogue metadata through controlled-vocabulary tagging and DCAT/GeoDCAT-AP-compatible representations. By operationalizing these mechanisms within real data space infrastructures, this work provides an early blueprint for integrating semantic technologies into the future CEADS federation.

## 2. Background

Data spaces have emerged as a paradigm for data sharing in which participants retain sovereignty over their data while enabling trust, traceability (often blockchain-enabled) and interoperability through shared governance, technical standards, and semantic agreements [9]. While data spaces can operate in totally free-for-users environments, they also play an important role in the emerging data economy, as they enable participants to monetize controlled access to their data assets [10]. Within this paradigm, CEADS envisions interoperability across multiple DSIs, each responsible for local data management while contributing to federation-level discovery and reuse. In this context, the DSSC Blueprint highlights semantic interoperability as a foundational requirement, emphasizing the role of shared data models, such as controlled vocabularies, ontologies, and metadata schemas, to ensure consistent, machine-interpretable data exchange across data spaces [3].

### 2.1. CEADS Use Case 2 and DSIs

CEADS operationalizes this vision through a set of eight domain-driven use cases that contribute heterogeneous datasets governed by distinct technical, legal, and business models. Among them, Use Case 2 – *Research and innovation data for improved AI algorithms performance*[6] – focuses on

[4]https://www.w3.org/TR/vocab-dcat-3/
[5]https://blueprint.dssc.eu/?pane=tools
[6]https://ceads.eu/pilots/research-and-innovation-data-for-improved-ai-algorithms-performance/

enhancing the quality, interoperability, and reusability of agricultural data to support trustworthy, semantically enabled high-performing artificial intelligence. This objective explicitly emphasizes the role of semantic technologies, including standardized metadata, controlled vocabularies, and ontology-based annotation, to contribute to data AI-readiness by improving data discoverability, integration across sources, and machine interpretability for AI-driven analytics and processing. The use case also establishes a connection with the European Open Science Cloud (EOSC), considered as the European data space for research and innovation, where semantic interoperability and the FAIR Principles are likewise central to enabling data sharing and reuse [11, 12]. Within this use case, three DSIs (AgrospAI, AgrifoodTEF, and DjustConnect[7]) collaborate to provide complementary infrastructures and datasets, collectively illustrating how semantic interoperability mechanisms can strengthen AI readiness across the CEADS federation.

AgrospAI and AgrifoodTEF are built upon the Pontus-X (pontus-x.eu) technology and infrastructure, which provides mechanisms for secure data sharing, governed access, and controlled execution of computations as well as a shared blockchain to record all the operations related to assets (*datasets*, *algorithms*, or *services*). A central architectural principle of Pontus-X is its Compute-to-Data (CtD) implementation (derived from the Ocean Protocol – oceanprotocol.com) in which algorithms are executed within a trusted environment where data reside, rather than transferring data to external processing environments [13, 14, 15]. In this model, dataset owners retain control over their assets by specifying which algorithms are authorized to run on their data. This approach preserves data sovereignty while still enabling analytics and service execution across organizational boundaries.

Within Pontus-X, assets are identified through a Decentralized IDentifier (DID) that resolves to a DID document (DDO), a JSON metadata record containing descriptive information, signatures, and access details.[8] These records are indexed in the Aquarius metadata catalogue, which is shared across all Pontus-X deployments, enabling consistent metadata access and discovery [16, 17]. These internal models support governance, access control, and compute orchestration, forming the operational backbone through which AgrospAI and AgrifoodTEF participate in the broader CEADS federation.

## 2.2. Catalogue-level standards for the agricultural data space

Achieving interoperability across DSIs requires more than secure exchange mechanisms; it also depends on shared data models, as framed in the DSSC Blueprint [18], or on semantic artefacts, as these resources are commonly referred to in the EOSC context [11]. These include vocabularies, ontologies, and metadata schemas that enable machine-interpretable discovery, comparison, and reuse of data assets. Ontologies and controlled vocabularies play a central role by providing stable concept identifiers, multilingual labels, and formally defined relationships that support semantic search, integration of heterogeneous datasets, and ontology-driven analytics [5, 6, 7].

At the catalogue level, this role is played by metadata schemas and application profiles. An application profile adapts a metadata schema to the needs of a particular community, domain, or interoperability scenario by selecting, constraining, and combining existing classes and properties, thereby enabling metadata records to be validated against the profile's requirements [19, 20]. DCAT, the Data Catalog Vocabulary, provides a general RDF vocabulary for describing catalogues, datasets, data services, and distributions [21]. It is widely used to support catalogue interoperability across data portals and is also adopted in data space reference architectures and connector-based implementations, including IDSA, Gaia-X, Eclipse Dataspace Components, and FIWARE [22]. DCAT-AP is the European application profile of DCAT, defining a harmonized way to describe datasets across European data portals [23]. GeoDCAT-AP further specializes this approach for geospatial metadata by aligning DCAT-based descriptions with spatial metadata requirements from INSPIRE and ISO 19139/19115 [24, 25, 26]. This makes GeoDCAT-AP particularly relevant for agricultural datasets, many of which contain georeferenced data.

[7] https://djustconnect.be/en

[8] An example of data asset on Pontus-X ecosystem accessed via Aquarius API: https://aquarius.pontus-x.eu/api/aquarius/assets/ddo/did:op:4d6f4ca585ad87372950da178689e2baf8b8e2668968765ca305bfb718af9cf3

In addition to GeoDCAT-AP, governance-aware interoperability in data spaces relies on complementary specifications such as ODRL[9] for usage policies and W3C Verifiable Credentials[10] for decentralized trust, as highlighted in the DSSC Blueprint [3]. While these aspects are not the primary focus of the present paper, aligning semantic enrichment and metadata exposure with this standards is essential for future CEADS federation.

### 2.3. Ontology repositories and the role of AgroPortal in CEADS

Semantic annotation in domain-specific data spaces depends on access to curated vocabulary services that provide stable, reusable vocabularies [5, 6]. The OntoPortal Alliance (ontoportal.org) offers a family of domain-oriented ontology repositories that support ontology hosting, browsing, search, and automated annotation services [27] both via a web application and through web APIs. OntoPortal has been widely adopted in the context of EOSC as for semantic artefact catalogues (i.e., vocabulary service) supporting FAIR and semantic interoperability requirements [28]. As such, OntoPortal provides a mature semantic infrastructure that can be reused across different sectoral data spaces.

AgroPortal (agroportal.eu) is the OntoPortal instance dedicated to the agrifood domain, providing a large collection of controlled vocabularies and ontologies relevant to crops, livestock, environment, and food systems [29]. Beyond simple vocabulary access, AgroPortal enables ontology-based semantic annotation via the Search or Annotator services that help enrich both data content and metadata descriptions. AgroPortal connectors have previously been developed to support ontology-based tagging of metadata records. For example, within the FAIR-IMPACT initiative, the INRAE Data Portal (an instance of Dataverse) implemented a metadata form widget incorporating a drop-down selector that enables dataset description through concepts drawn directly from reference ontologies [30].

These characteristics make AgroPortal well aligned with CEADS requirements. By providing curated semantic artefacts and machine-actionable annotation services, AgroPortal can support consistent thematic classification, multilingual interoperability, and ontology-driven discovery across DSIs.

Among the semantic artefacts available through AgroPortal, AGROVOC plays a central role for CEADS. As FAO's multilingual thesaurus for food and agriculture [31], it provides stable concept identifiers, labels, synonyms, and semantic relationships. In CEADS, AGROVOC is being adopted as the domain-specific controlled vocabulary for tagging data assets and is therefore used as the default reference vocabulary in the demonstrations presented in this paper.

## 3. Related work

Recent studies on semantics in data spaces have emphasized that interoperability requires more than technical connectivity between participants. Hauff et al. [32] argue that semantics provide an infrastructure-independent foundation for FAIR data management in data spaces, supporting requirements such as self-description, metadata, catalogues, standardization, validation, and enrichment. Similarly, David et al. [33] discuss vocabulary hubs as key components of semantic layers in data spaces, responsible for publishing and governing shared vocabularies, and propose their federation to support vocabulary sovereignty, governance, findability, and reuse across data spaces. Other work, such as the Semantic Data Link framework [34], addresses the practical difficulty of applying semantic technologies by providing more user-friendly mechanisms for creating interoperable descriptions while reusing existing vocabularies and generating machine-actionable artefacts. These studies frame semantic interoperability as a practical infrastructure concern rather than only a modeling problem.

A second line of related work concerns the application of domain-specific metadata application profiles for European data spaces and data portals. In the European Health Data Space, HealthDCAT-AP extends DCAT-AP with health-specific metadata elements, controlled vocabularies, refined cardinalities, and validation mechanisms to support cross-border discovery, access, and reuse of electronic health

[9] https://www.w3.org/TR/odrl-model/
[10] https://www.w3.org/TR/vc-data-model-2.0/

data [35]. In the mobility domain, mobilityDCAT-AP was developed as a metadata specification for National Access Points and mobility data portals, with the aim of harmonizing metadata descriptions and improving cross-portal discovery. Although originally centered on the federation of mobility data portals, it is now also being promoted in the context of the European Mobility Data Space, notably through deployEMDS, as a reference metadata profile for operational mobility data spaces and their interlinking layer [36]. In the cultural domain, the Culture Information Model proposes an extensible semantic information model and an AP-first methodology for developing application profiles, ontologies, and controlled vocabularies for domain-specific data spaces [37]. These initiatives show that sectoral data spaces often require more than generic DCAT-AP: they need domain-specific constraints, controlled vocabularies, validation mechanisms, governance processes, and adoption support.

Other studies focus on the operational connection between existing data infrastructures and data space catalogues. In the Green Deal Data Space context, Masó et al. [26] address a closely related problem: how to make geospatial data assets discoverable and accessible by combining established geospatial metadata infrastructures with data space technologies. Their architecture connects GeoNetwork-based ISO 19115/19139 metadata catalogues, cloud storage, processing services, and data space connectors, while recognizing DCAT and GeoDCAT-AP as key bridges between geospatial cataloguing practices and data space catalogues. This work is particularly relevant to CEADS because we face a similar need to reconcile INSPIRE/ISO-based metadata with DCAT/GeoDCAT-AP representations for federated discovery. The connection is also practical: several actors involved in Green Deal Data Space activities are also contributing to CEADS, bringing experience into the agricultural domain. Morejón et al. [22] complement this perspective by arguing that high-level catalogue metadata may be insufficient for discovery in data spaces, and by proposing a DCAT extension with content-level metadata, such as data dictionaries, fields, and representative sample values, mapped to data ecosystem catalogues such as CKAN, EDC, and IDSA.

Compared with these studies, the present paper does not propose a complete agriculture-specific DCAT-AP extension or a new semantic modeling methodology for CEADS. Instead, it focuses on the operational integration of semantic enrichment mechanisms within Pontus-X-based DSIs participating in CEADS. The contribution is therefore implementation-oriented: controlled-vocabulary selection at publication time, ontology-based annotation in a compute-to-data workflow, and DCAT-AP/GeoDCAT-AP-compatible metadata exposure from native Pontus-X records.

## 4. Contributions: Ways of integrating semantics into CEADS

### 4.1. Providing semantic annotations for datasets using AgroAnnotator

AgroAnnotator is a CtD algorithm, written in Python 3.11 that performs ontology-based annotation of agricultural documents/datasets based on the AgroPortal Annotator service (https://agroportal.eu/annotator)[11] inside the secure AgrospAI execution environment, as shown in Fig. 1. The algorithm and its documentation are available on GitHub [1].

The workflow begins when a user launches a compute job by selecting an input dataset (document) and the AgroAnnotator algorithm in AgrospAI. The algorithm supports heterogeneous input formats, including PDF, HTML, DOCX, and plain text (e.g., TXT and Markdown), and converts them into a plain-text representation through format-specific extraction procedures. For PDF files, text is extracted page by page using the PdfReader class from the PyPDF2 library, retaining only non-empty textual content and concatenating pages with paragraph-level separation. For DOCX documents, the algorithm iterates through paragraph elements and extracts non-empty paragraph text, preserving their sequential order. HTML inputs can be parsed using BeautifulSoup library, which allows removing non-content elements such as script, style, and noscript tags, after which the textual content is extracted while preserving basic structural cues through line breaks. If the BeautifulSoup library is unavailable in

[11]The AgroPortal Annotator is the AgroPortal specific instance of the Annotator service incorporated in the OntoPortal technology presented and described by previous works [38, 39].

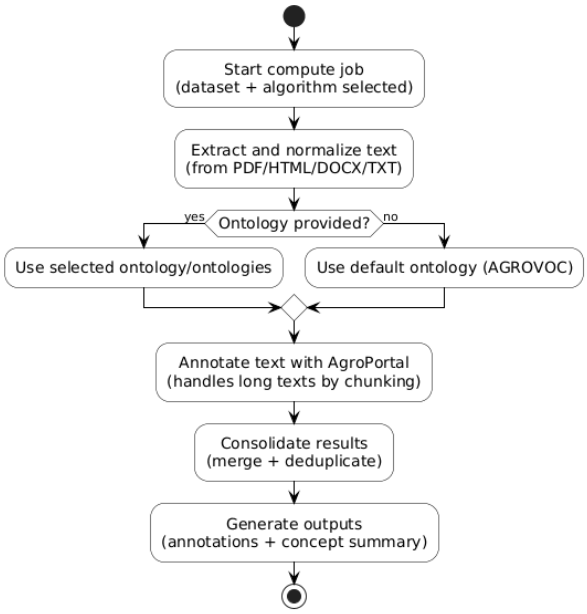



**Figure 1:** AgroAnnotator workflow.

the run time environment, the algorithm includes a fallback parser that is used to approximate this behaviour by selectively handling block-level elements and ignoring non-visible content. Plain text and Markdown files are read directly without structural transformation. In cases where file extensions are missing or ambiguous, the algorithm performs lightweight content sniffing based on file signatures (e.g., %PDF- headers or HTML tags) to infer the appropriate parsing strategy.

Following extraction, all inputs undergo a lightweight normalization step to ensure consistency across formats. This includes standardizing line endings, collapsing redundant whitespace, and reducing excessive blank lines while preserving basic paragraph structure. The resulting normalized text constitutes a unified representation that abstracts away format-specific differences while retaining the core textual content of the document. Rather than performing explicit structural or semantic segmentation (e.g., identifying titles, abstracts, or sections), the algorithm adopts a format-agnostic approach in which the entire textual content is considered for downstream processing. For long documents, AgroAnnotator partitions the text into manageable chunks (max. 8000 characters).

Within AgrospAI, users may list one or more ontologies to use for annotation with their acronyms corresponding to their identifier in AgroPortal. If none are supplied, the algorithm defaults to AGROVOC only. The normalized text is then submitted to the AgroPortal Annotator API together with the ontologies selected for annotation. The AgroPortal service returns structured JSON annotations describing detected concepts for each chunk. The Annotator is called with default parameters (i.e., whole word only, select ontologies), and without hierarchy or mapping-based semantic expansion.

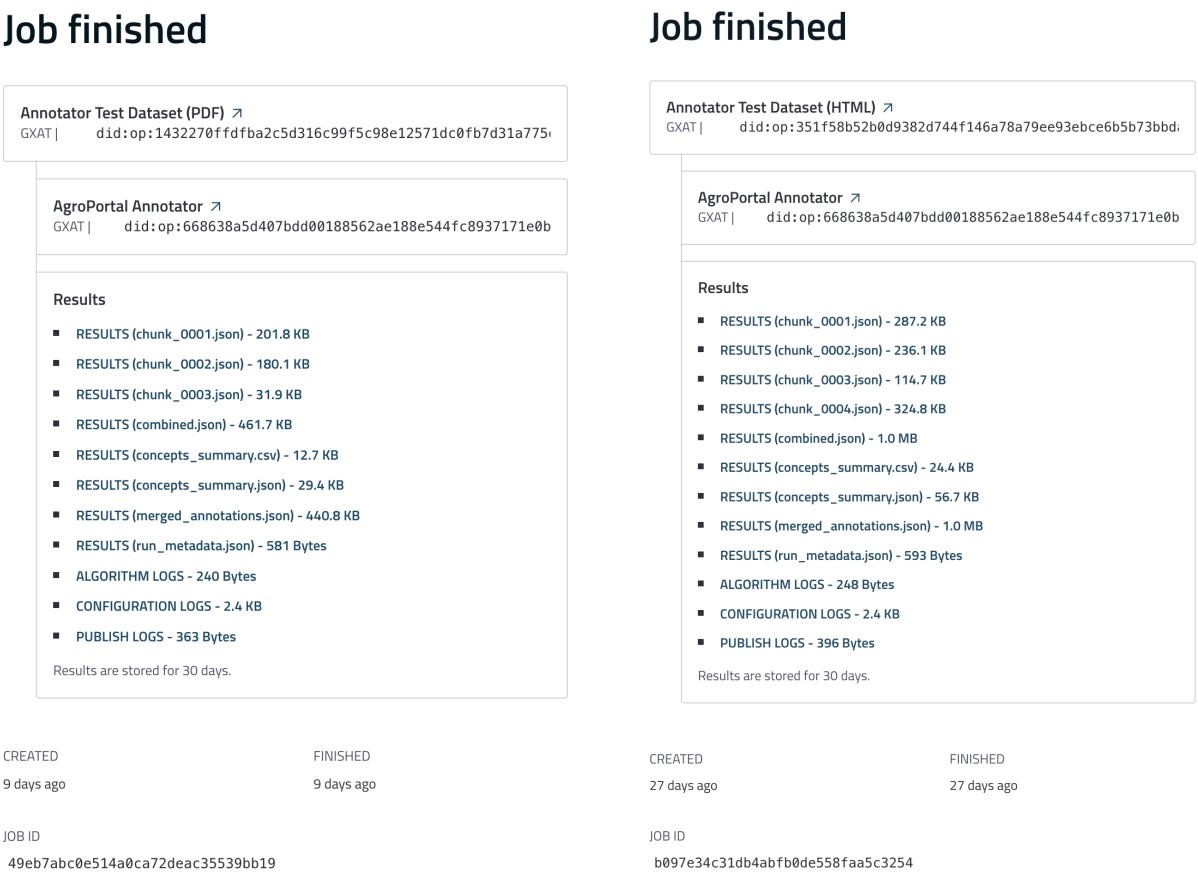
Finally, AgroAnnotator consolidates chunk-level results into a single output by merging and deduplicating annotations across chunks. It generates a unique final response document including raw chunk annotations, merged document-level concept annotations, and aggregated concept summaries (which are scores including the number of unique concepts matched, and the number of times each concept appears in the text) in both JSON and CSV formats.

The intrinsic quality of the annotation pipeline underlying the AgroPortal Annotator has been evaluated in previous work [38, 39]. These evaluations show that the service generally achieves good precision, although recall depends on factors such as ontology coverage, lexical richness, and the characteristics of the input text. A detailed evaluation of annotation quality is therefore beyond the scope of this paper. Instead, our focus is on integrating this existing semantic annotation service into a Pontus-X compute-to-data environment. This integration also benefits from the genericity of AgroPortal: any ontology available in the repository can be used by the Annotator without workflow-

specific redevelopment, and future improvements to the Annotator can be inherited by AgroAnnotator without substantial changes to the surrounding data space workflow.

To test the prototype, two CtD jobs were executed on AgrospAI to validate the end-to-end workflow with different input formats. First, AgroAnnotator was published as an algorithm asset in AgrospAI and assigned a persistent DID identifier [2]. Two datasets were then published using the INRAE credential[12] to serve as inputs for the algorithm: Annotator Test Dataset (HTML) [3] and Annotator Test Dataset (PDF) [4]. These example documents can be tested by any user identified in AgrospAI. At the algorithm level, no restrictions are imposed regarding who can execute the workflow or which datasets can be used as input. Instead, access control and usage permissions are managed by the underlying Pontus-X dataspace infrastructure, which governs dataset visibility and execution rights through its credential and policy mechanisms.

In both runs, AgroAnnotator executed successfully inside the AgrospAI compute environment, demonstrating that the same algorithm can process heterogeneous document types while preserving the CtD constraint (i.e., the source documents remain under the control of the provider, and the consumer only gets access to the algorithm's output). Each job produced the expected set of outputs. The results are shown in Fig. 2.




(a) Compute job execution output (PDF) (b) Compute job execution output (HTML)

**Figure 2:** Screenshots from the compute job execution output in AgrospAI: (A) PDF input, (B) HTML input.

Listing 1 shows a representative excerpt of the semantic annotations generated by AgroAnnotator. The output is structured as a JSON array where each element corresponds to a concept identified in a reference ontology. For each detected concept, the field `annotatedClass` provides the ontology identifier (URI) of the matched class – in this case from the AGROVOC vocabulary – while the `annotations` field lists the occurrences of the concept detected in the input text. Each annotation records the matched text

[12] https://portal.agrospai.udl.cat/profile/0x45cce053ab6Cc02dFF4B24222D769eF90fBEDfF5

string (`text`), the type of match (`matchType`, e.g., preferred label or synonym), and the character offsets (`from`, `to`) indicating the position of the match within the original text.

Listing 1: Excerpt of AgroAnnotator semantic annotations linking text mentions to AGROVOC concepts


```
[
 {"annotatedClass":{"@id":"http://aims.fao.org/aos/agrovoc/c_f888df97","@type":"owl:Class"},
  "annotations":[
    {"text":"CITIZEN SCIENCE","matchType":"PREF","from":186,"to":200},
    {"text":"CITIZEN SCIENCE","matchType":"PREF","from":676,"to":690}
  ]},
 {"annotatedClass":{"@id":"http://aims.fao.org/aos/agrovoc/c_e89c7744","@type":"owl:Class"},
  "annotations":[{"text":"FOOD FRAUD","matchType":"PREF","from":435,"to":444}]},
 {"annotatedClass":{"@id":"http://aims.fao.org/aos/agrovoc/c_4911","@type":"owl:Class"},
  "annotations":[{"text":"MONITORING","matchType":"PREF","from":631,"to":640}]},
 ...
]
```


## 4.2. Tagging data asset metadata with controlled vocabularies

As a second contribution, we present two publication-time prototypes for enriching Pontus-X asset metadata with controlled vocabulary terms in AgrospAI and AgriFoodTEF. Unlike AgroAnnotator, which annotates dataset content through a compute-to-data algorithm, these prototypes operate at the metadata layer during asset publication or editing. Both complement free-form tags with controlled terms, while differing in implementation: AgrospAI uses AgroPortal as a generic vocabulary service, whereas AgriFoodTEF connects directly to the AGROVOC API. Work is ongoing to harmonize these implementations.

### 4.2.1. Integration between AgrospAI and AgroPortal

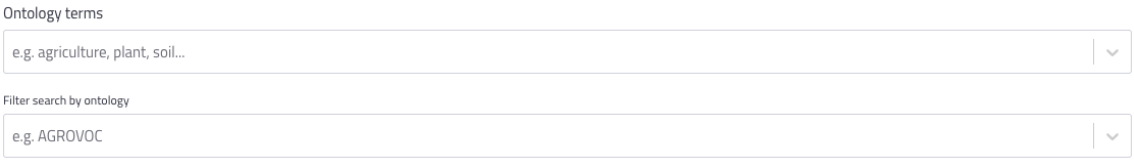
The first prototype consists of an integration between AgrospAI and AgroPortal that enables users to tag assets with controlled terms during the asset publication workflow. Pontus-X already supports free-form keywords (tags), which can lead to inconsistency due to spelling variants, synonyms, and multilingual drift. The integration complements this folksonomy with controlled terms explicitly taken from a selected referenced vocabulary served by AgroPortal, enabling more consistent and interoperable descriptions of datasets, algorithms, and services.

The integration is implemented in the AgrospAI publication workflow as a controlled-term selector using the AgroPortal Auto-complete Widget, as shown in Fig. 3. Users can search for concepts across AgroPortal-hosted semantic artefacts, optionally filtering the controlled vocabularies used for suggestions. Selected terms are stored as resolvable URIs with their preferred labels. This implementation is already deployed and available to users.[13]

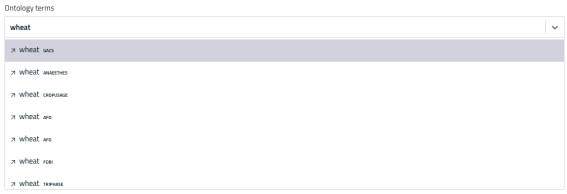
Although CEADS currently promotes AGROVOC as the domain-specific controlled vocabulary for tagging agricultural data assets, this choice may evolve as the data space matures and additional vocabularies are adopted for specific metadata requirements. For this reason, the connector was designed as a generic AgroPortal-based component rather than as an AGROVOC-specific integration. This design allows the same mechanism to support other controlled vocabularies available through AgroPortal, without requiring changes to the publication workflow.

For instance, future CEADS metadata workflows will require the use of the High-Value Dataset (HVD) Categories vocabulary,[14] a controlled vocabulary published by the Publications Office of the European Union through EU Vocabularies. This vocabulary supports the classification of HVDs according to the categories defined in Commission Implementing Regulation (EU) 2023/138. Since the HVD Categories vocabulary is also available in AgroPortal,[15] the same connector used to select AGROVOC concepts can also be used to select HVD category concepts. The connector could easily be extended to work with the OntoPortal Federation's API (work in progress) [40]. In this sense, the connector provides

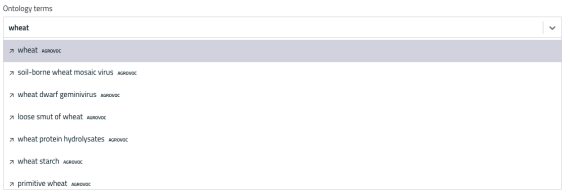
[13] https://portal.agrospai.udl.cat/publish/1
[14] http://data.europa.eu/bna/asd487ae75
[15] https://agroportal.eu/ontologies/HVDC




(a) Ontology term selector fields in the publication workflow. The list of available ontologies and controlled vocabularies is retrieved from AgroPortal, and each ontology acronym must match the acronym used in AgroPortal.




(b) Searching "wheat" without ontology filtering shows matches across multiple ontologies.




(c) Searching "wheat" with AGROVOC selected restricts suggestions to that controlled vocabulary.

**Figure 3:** AgroPortal integration in AgrospAI's user interface for selecting controlled terms.

a reusable controlled-vocabulary selection service that can support both current agricultural-domain tagging needs and future regulatory or cross-domain metadata requirements in CEADS.

To preserve compatibility with the Pontus-X DDO metadata, controlled terms are not stored in the existing DDO `categories` field (which is constrained to an array of strings). Instead, they are included under the DDO field `additionalInformation`, which is intended for metadata extensions. Following the DCAT recommendation for dataset classification, the connector represents the selected concepts as `dcat:theme` entries, each modeled as a `skos:Concept`, as shown in Listing 2. The complete metadata record can be accessed through the public AgrospAI Aquarius API endpoint [5].

Listing 2: Example of semantic annotations in the `additionalInformation` field

```
additionalInformation:{
 @context:{...},
 dcat:theme:[
  {"@id":"http://aims.fao.org/aos/agrovoc/c_541","@type":"skos:Concept",
   "skos:prefLabel":{"@value":"apples","@language":"en"}},
  {"@id":"http://aims.fao.org/aos/agrovoc/c_3131","@type":"skos:Concept",
   "skos:prefLabel":{"@value":"fruits","@language":"en"}},
  {"@id":"http://aims.fao.org/aos/agrovoc/c_92363","@type":"skos:Concept",
   "skos:prefLabel":{"@value":"precision agriculture","@language":"en"}}
 ]
}
```

This prototype demonstrates how AgroPortal can act as a generic controlled-vocabulary connector for Pontus-X-based publication workflows. The documentation of this prototype is openly available from GitHub [6].

### 4.2.2. Integration between AgriFoodTEF and AGROVOC

The second prototype was developed in the Fondazione Bruno Kessler Data Space,[16] a Pontus-X-based environment that serves as a testbed for the AgriFoodTEF DSI. It addresses the same metadata enrichment objective through an AGROVOC-specific implementation, which connects directly to the AGROVOC Skosmos REST API[17] to support concept search and selection during asset publication.

In the publication workflow,[18] a new tab named *AgriMetadata* was added to the asset publication form. This tab allows users to provide GeoDCAT-AP-relevant metadata, such as geographic and temporal

[16] https://dataspace.fbk.eu
[17] https://www.fao.org/agrovoc/machine-use
[18] https://dataspace.fbk.eu/publish/

coverage (discussed in Section 4.3), and to semantically tag assets with AGROVOC concepts. As shown in Fig. 4, AGROVOC concepts can be searched through an autocomplete field integrated into the publication interface. When the user enters a term, the interface retrieves matching AGROVOC concepts and allows the user to select the most appropriate ones for describing the asset. Each selected concept is stored using its AGROVOC URI together with its preferred label, thereby complementing free-form keywords with stable, machine-actionable concept identifiers.

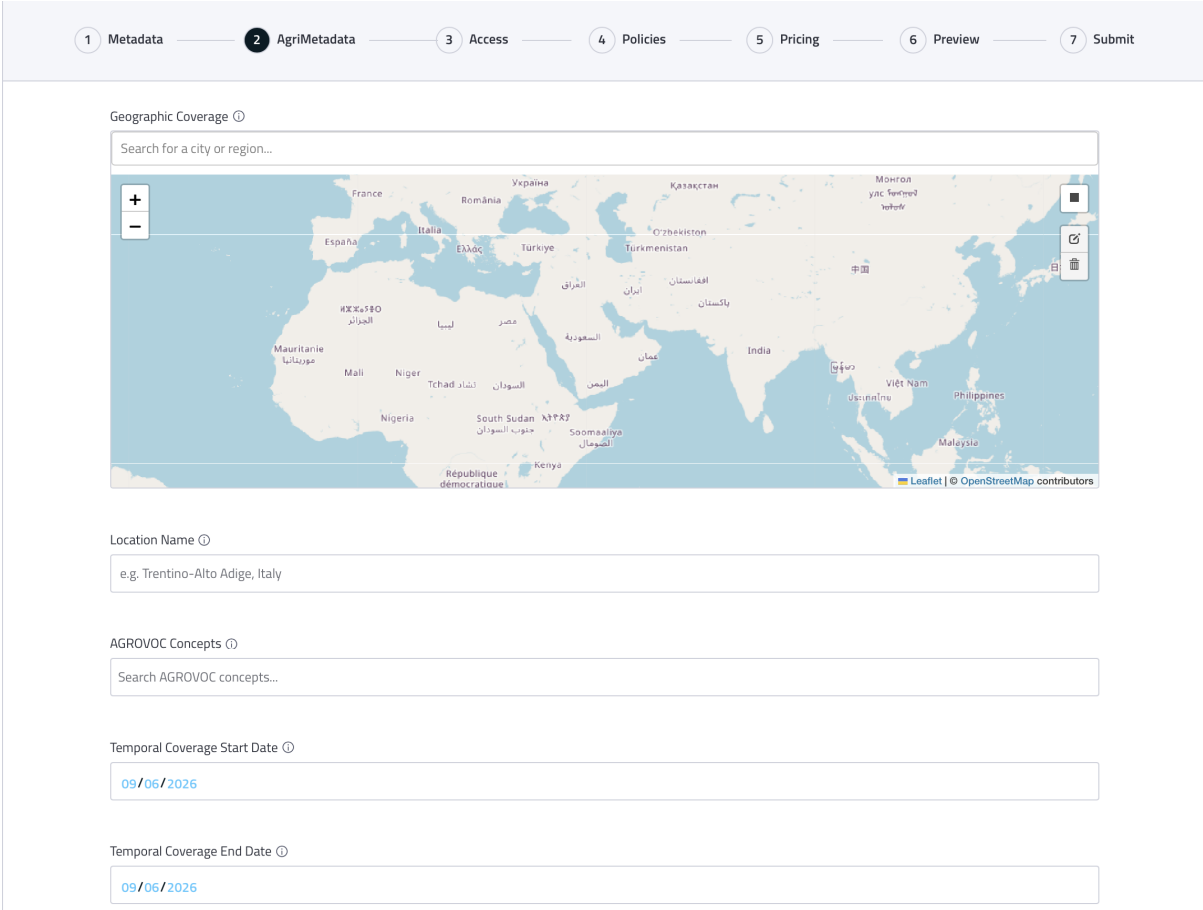



**Figure 4:** AgriMetadata tab in the AgriFoodTEF publication workflow, including fields for GeoDCAT-AP-relevant metadata, such as geographic and temporal coverage, and an autocomplete field for selecting AGROVOC concepts.

This implementation demonstrates a more direct and lightweight approach to semantic tagging than the AgroPortal-based connector. It is well-suited to contexts in which AGROVOC has already been selected as the reference vocabulary for describing the agricultural assets. At the same time, it is less generic than the AgroPortal integration, because it is tied specifically to AGROVOC and does not directly support switching to other controlled vocabularies.

After the asset is published in the catalogue, the selected AGROVOC concepts are displayed in the asset description interface, as shown in Fig. 5. This integration makes the semantic tags visible to users and will support semantic search in the future.

Similar to the AgrospAI prototype, this implementation records controlled vocabulary terms by embedding the selected concepts, including their URIs and preferred labels, inside `additionalInformation` (see Listing 2). The full JSON-LD metadata record corresponding to the example shown in Fig. 5 is available through the Aquarius API [8].

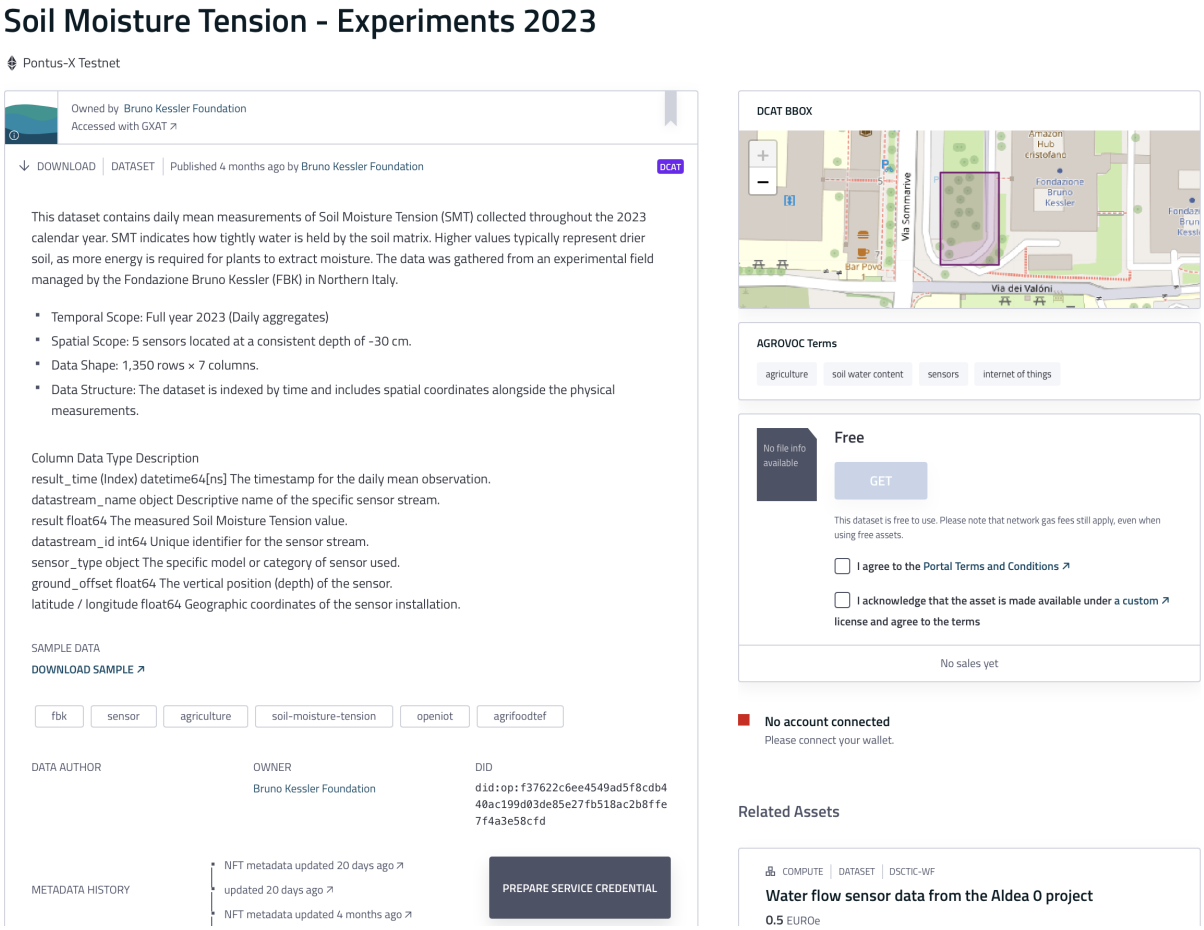



**Figure 5:** Asset description in the AgriFoodTEF user interface, showing the GeoDCAT bounding box and the selected AGROVOC terms displayed in the metadata panel on the right side.

## 4.3. Generating GeoDCAT-AP compliant metadata

As a third contribution, we present catalogue-level services that expose Pontus-X records as DCAT or GeoDCAT-AP-compatible metadata. In Pontus-X, assets are indexed in Aquarius using DDO, which supports the internal operation of the ecosystem but is not directly aligned with EU catalogue interoperability recommendations based on DCAT and GeoDCAT-AP. The implementations described below bridge this gap by transforming or exposing Pontus-X metadata as standards-aligned JSON-LD.

The implementations described in this section bridge this gap by transforming or exposing Pontus-X metadata as standards-aligned JSON-LD. We first present the GeoDCAT-AP wrapper implemented for AgrospAI, which maps individual DDO records into GeoDCAT-AP 3.0-compliant representations. We then present the related implementation developed in the Fondazione Bruno Kessler Data Space, used as a testbed for the AgriFoodTEF Data Space, where additional agricultural metadata are collected during publication and exposed through a DCAT-compatible catalogue endpoint.

### 4.3.1. GeoDCAT-AP wrapper for AgrospAI

In AgrospAI, the GeoDCAT-AP wrapper is implemented as a catalogue-level service that transforms Pontus-X DDO records into GeoDCAT-AP 3.0-compliant JSON-LD representations. The wrapper follows a loss-minimizing transformation strategy. First, it retrieves the original DDO record[19] from Aquarius and maps fields with clear semantic equivalence into their GeoDCAT-AP counterparts (e.g., title, description, license, dates, creator/publisher information, keywords). The mapping script available from

[19]Example of a DDO metadata record: https://aquarius.pontus-x.eu/api/aquarius/assets/ddo/did:op:4d6f4ca585ad87372950da178689e2baf8b8e2668968765ca305bfb718af9cf3

GitHub [9]. Second, any DDO attributes that do not have an unambiguous mapping are preserved under a dedicated `pontus-x` namespace to avoid information loss and to keep traceability back to the source representation. Third, and importantly for CEADS requirements, the wrapper reuses structured semantic metadata stored in the `additionalInformation` field and maps it to the corresponding GeoDCAT-AP properties. For example, spatial information, such as bounding boxes and place labels, is exposed through `dct:spatial`, while thematic concepts selected from controlled vocabularies are exposed through dcat:theme as `skos:Concept` resources.[20] This mechanism allows AgrospAI to combine: (i) the core Pontus-X metadata, (ii) publisher-provided free-text tags/keywords, and (iii) structured semantic tagging (such as AgroPortal-derived themes and geospatial annotations) into a single standards-aligned description suitable for external catalogues and discovery services.

The resulting JSON-LD output has been validated against the GeoDCAT-AP 3.0 SHACL shapes[21] using the SHACL Validator tool [41] from Interoperable Europe without errors or warnings, demonstrating that Pontus-X catalogue entries can be exported in a standards-compliant form. This wrapper therefore operationalizes catalogue-level semantics in AgrospAI by enabling direct consumption of their metadata in a GeoDCAT-AP compliant format, while maintaining compatibility with the Pontus-X internal schema and preserving the richer, platform-specific attributes required for CtD execution. It is worth mentioning that this wrapper works for any data space that uses the Pontus-X technology.

#### 4.3.2. DCAT-compatible metadata exposure in AgriFoodTEF

The *AgriMetadata* tab introduced in Section 4.2.2 (Fig. 4) also collects GeoDCAT-AP-oriented metadata, namely spatial and temporal coverage, in addition to AGROVOC concepts. Spatial coverage is captured through a map-based interface that allows the publisher to define a geographic bounding box and associate it with a human-readable location name. Temporal coverage is captured through start and end dates describing the temporal extent of the asset. Together with the AGROVOC concepts, these fields are stored in the DDO under `additionalInformation` using the same representation as the AgrospAI prototype (Listing 2). This design preserves compatibility with the Pontus-X DDO while making the added metadata available for downstream transformation into DCAT-compatible JSON-LD.

As an optional alternative to entering these fields manually, the prototype also offers a metadata-import helper, currently being tested, that pre-populates them from an external catalogue record given its URI. Internally, this helper relies on a small microservice, implemented in Python, that converts the external record into the expected DDO structure, including the additional metadata described above. This feature targets well-known catalogue formats and platforms, including CKAN, GeoNetwork, and STAC catalogues, as shown in Fig. 6.

To expose the enriched metadata in DCAT-compatible form, the captured bounding box is rendered on a map in the asset description interface (Fig. 5), making the spatial descriptor visible to users while preserving its machine-actionable representation. The same metadata is also served programmatically through a REST endpoint that returns the catalogue as DCAT-compatible JSON-LD,[22] serialized as a `dcat:Catalog` whose published assets are `dcat:Resource` entries. In this representation, Pontus-X metadata is mapped to DCAT properties, AGROVOC concepts are exposed as thematic descriptors, and spatial and temporal coverage through the corresponding GeoDCAT-AP-compatible properties.

The output of this endpoint was tested using the DCAT validator,[23] resulting only in minor warnings. In the AgriFoodTEF interface, validation results are also exposed to users through a DCAT badge: when an asset metadata record passes validation without critical warnings, the asset is marked as DCAT-compliant, as shown in Fig 5. Beyond this specific implementation, such badges illustrate a possible direction for Pontus-X-based data spaces: metadata validation results could be made visible at the asset level to indicate whether a dataset conforms to a given standard or application profile.

[20] Example of a DDO metadata record mapped to GeoDCAT-AP: https://dcat.agrospai.udl.cat/api/aquarius/assets/ddo/did:op:4d6f4ca585ad87372950da178689e2baf8b8e2668968765ca305bfb718af9cf3

[21] https://semiceu.github.io/GeoDCAT-AP/releases/3.0.0/shacl/geodcat-ap-SHACL.ttl

[22] AgriFoodTEF DCAT catalogue endpoint: https://dataspace.fbk.eu/api/dcat/catalog

[23] http://www.dcat.be/validator/

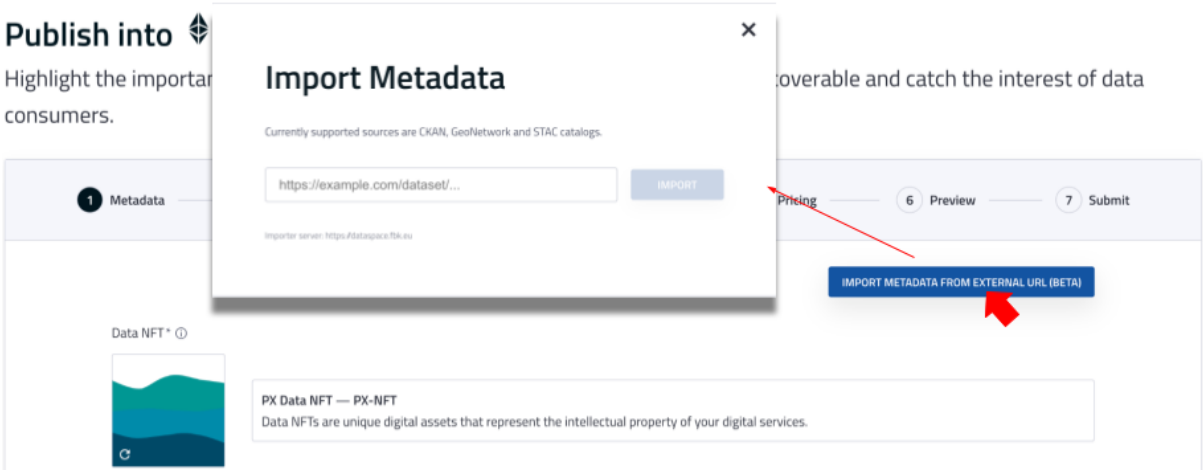



**Figure 6:** Prototype feature for importing metadata from an external catalogue record using its URL.

## 5. Discussion

The AgrospAI and AgriFoodTEF implementations show two complementary strategies for introducing catalogue-level semantics into Pontus-X-based systems. Together, these approaches show that Pontus-X-native metadata can be progressively aligned with DCAT and GeoDCAT-AP without replacing the underlying DDO model.

The two prototypes reveal complementary strengths that we are currently working to converge into a unified and harmonized implementation for potential contribution to the upstream Pontus-X codebase. AgrospAI contributes a generic AgroPortal-based connector, independent from a single controlled vocabulary, together with a GeoDCAT-AP wrapper validated against available SHACL validators. AgriFoodTEF contributes stronger user-interface integration, including publication-time fields for AGROVOC concepts, spatial and temporal coverage, DCAT badges, map-based bounding box visualization, and display of selected AGROVOC terms. We are currently working in converging these two implementations so that they can be contributed to the upstream Pontus-X codebase.

At this stage, our validation is primarily functional. We verified end-to-end execution in the AgrospAI environment, including publication, compute job execution, and output generation, and confirmed that the GeoDCAT-AP wrapper produces GeoDCAT-AP 3.0-compliant metadata via SHACL validation. For AgriFoodTEF, we verified that additional publication-time metadata can be stored in the DDO, displayed in the user interface, and exposed through a DCAT-compatible catalogue endpoint. However, further validation is still needed to assess metadata quality, usability for publishers, consistency of concept selection, and the practical impact of the enriched metadata on dataset discovery across DSIs.

Ontology-based annotation and controlled vocabulary indexing are well-established techniques in the semantic web and information retrieval communities, and their benefits for improving data discovery, interoperability, and semantic integration have been widely demonstrated in previous works (e.g., [42]). Rather than proposing a new annotation method, the contribution presented here focuses on bringing these established semantic enrichment practices into real Pontus-X-based data space environments and evaluating how they can be operationalized within CEADS.

Regarding AgroAnnotator, we plan to extend its functionality beyond document annotation toward the structured semantic harmonization of dataset labels. In particular, the algorithm could ingest datasets that comply with community-defined schemas or standards, such as predefined columns and value conventions, and automatically generate annotations linking dataset elements to selected reference ontologies. For instance, AgroAnnotator could natively support the MIAPPE standard [43] for plant phenotyping datasets, enabling the automatic alignment of dataset fields and values with relevant ontology concepts.

A current limitation of AgroAnnotator is that text chunks are sent to the AgroPortal Annotator API

for concept recognition, meaning that portions of the data leave the compute environment. This is not fully aligned with data space principles. Future work will explore deploying annotation services inside the DSI environment or restricting external calls to non-sensitive metadata fields.

## 6. Conclusions

This paper presented a set of complementary prototypes for strengthening semantic interoperability in Pontus-X-based DSIs participating in CEADS. Rather than replacing existing DSI workflows, the proposed components introduce semantics at different points of the asset lifecycle: during compute, through ontology-based annotation with AgroAnnotator; at publication time, through controlled-vocabulary enrichment in AgrospAI and AgriFoodTEF; and at catalogue level, through DCAT/GeoDCAT-AP-compatible metadata exposure. Together, these implementations show that semantic enrichment can be embedded incrementally into existing Pontus-X infrastructures while preserving compatibility with the native DDO model.

Beyond catalogue interoperability alone, these contributions also support the broader CEADS Use Case 2 objective of making agricultural data more AI-ready. By bringing semantics into CEADS, the proposed components make agricultural data easier to discover, compare, interpret, and combine across DSIs. This is particularly important for AI workflows, where the usefulness of data depends not only on technical access, but also on the availability of contextual, machine-actionable metadata that supports reuse.

From a FAIR perspective, the prototypes contribute especially to findability, interoperability, and reusability. Controlled-vocabulary enrichment improves findability by enabling concept-based discovery beyond free-text keywords. DCAT/GeoDCAT-AP-compatible metadata exposure supports interoperability by aligning Pontus-X records with widely adopted catalogue standards and application profiles. Ontology-based annotation contributes to reusability by adding semantic context to data assets, helping users and machines assess their relevance for downstream analysis. In this sense, semantic enrichment acts as a bridge between FAIR data management and AI-readiness in agricultural data spaces.

## Data and Code Availability


1. AgroPortal. (Mar 10, 2026). AgroAnnotator for AgrospAI [Computer software]. GitHub. https://github.com/agroportal/agroannotator
2. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). (2026, Feb 9). AgroPortal Annotator - AgrospAI [Algorithm asset]. AgrospAI. DID: https://portal.agrospai.udl.cat/asset/did:op:668638a5d407bdd00188562ae188e544fc8937171e0b1b69f130fc4de413d178
3. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). (2026, Feb 9). Annotator Test Dataset (HTML) [Dataset]. AgrospAI. DID: https://portal.agrospai.udl.cat/asset/did:op:351f58b52b0d9382d744f146a78a79ee93ebce6b5b73bbda2a65af594bb36a54
4. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). (2026, Feb 9). Annotator Test Dataset (PDF) [Dataset]. AgrospAI. DID: https://portal.agrospai.udl.cat/asset/did:op:1432270ffdfba2c5d316c99f5c98e12571dc0fb7d31a775e9644319227b85ef2
5. Universitat de Lleida. (2026, Jan 9). AmodalAppleSize_RGB-D dataset: Apple tree pictures (20) [Metadata record]. Aquarius API. DID: https://aquarius.pontus-x.eu/api/aquarius/assets/ddo/did:op:4d6f4ca585ad87372950da178689e2baf8b8e2668968765ca305bfb718af9cf3
6. Universitat de Lleida. AgrospAI Data Space [GitHub repository]. https://github.com/AgrospAI/mvg-portal
7. AgrospAI script to publish an asset with GeoDCAT metadata. https://github.com/AgrospAI/pontus-x_cli/blob/main/samples/publish/data/index.ts

8. Bruno Kessler Foundation. (2026, Feb 9). Soil Moisture Tension - Experiments 2023 [Metadata record]. Aquarius API. DID: https://aquarius.pontus-x.eu/api/aquarius/assets/ddo/did:op:f37622c6ee4549ad5f8cdb440ac199d03de85e27fb518ac2b8ffe7f4a3e58cfd
9. AgrospAI. (2026, Jan 14). dcat-ddo-wrapper.py [Source code]. GitHub. https://github.com/AgrospAI/dcat-ddo/blob/4055ba37cc995cbfd1ec2a8705e0edbe5cf47717/dcat-ddo-wrapper.py

## Acknowledgments

AgrospAI, the Agrifood Data Space Demonstrator for Sovereign Data-Sharing and Artificial Intelligence Services, is funded by the Spanish Ministry for Digital Transformation and the Civil Service under grant TSI-100120-2024-28. This work has received funding, in part, from the European Union's Digital Europe Programme under grant agreement No. 101195295.

## Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5.2 in order to: Grammar, spelling check, text flow, and to generate the PlantUML code for Fig. 1. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

[1] European Commission, Common European data spaces, 2025. URL: https://digital-strategy.ec.europa.eu/en/policies/data-spaces.
[2] International Data Spaces Association, Dataspace protocol 2024-1, 2024. URL: https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol.
[3] Data Spaces Support Centre, Data spaces blueprint v2.0, 2025. URL: https://dssc.eu/space/BVE2/1071251457/Data+Spaces+Blueprint+v2.0+-+Home.
[4] P. Colpaert, Dataspace challenges: Draft Community Group Report, 9 February 2026, Draft Community Group Report, W3C Community Group on Dataspaces, 2026. URL: https://w3c-cg.github.io/dataspaces/.
[5] M. Andresel, V. Siska, R. David, S. Schlarb, A. Weißenfeld, Adapting ontology-based data access for data spaces, in: J. Theissen-Lipp, P. Colpaert, S. K. Sowe, E. Curry, S. Decker (Eds.), The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 3705, CEUR Workshop Proceedings, 2024. URL: https://ceur-ws.org/Vol-3705/paper06.pdf.
[6] M. Hauff, L. M. Comet, P. Moosmann, C. Lange, I. Chrysakis, J. Theissen-Lipp, Fairness in dataspaces: The role of semantics for data management, in: J. Theissen-Lipp, P. Colpaert, S. K. Sowe, E. Curry, S. Decker (Eds.), The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 3705, CEUR Workshop Proceedings, 2024. URL: https://ceur-ws.org/Vol-3705/paper01.pdf.
[7] M. Stäbler, P. Moosmann, P. Dittmer, D. Wang, F. Köster, C. Lange, Semantic data link: Bridging domain-specific needs with universal and interoperable semantic models, in: J. Theissen-Lipp, P. Colpaert, S. K. Sowe, E. Curry, S. Decker (Eds.), The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 3705, CEUR Workshop Proceedings, 2024. URL: https://ceur-ws.org/Vol-3705/short13.pdf.
[8] F. Kirstein, M. Gysel, piveau-X: A compliance-focused semantic web-based catalog for data spaces, in: J. Theissen-Lipp, P. Colpaert, A. Pomp, E. Curry, S. Decker (Eds.), The Third International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 4007, CEUR Workshop Proceedings, 2025. URL: https://ceur-ws.org/Vol-4007/05short.pdf.

[9] A. Kovach, L. Montalvillo, J. Lanza, P. Sotres, A. Urbieta, Understanding data spaces: A systematic mapping study of foundations, technical building blocks, and sectoral adoption, Computer Science Review 59 (2026) 100819. doi:10.1016/j.cosrev.2025.100819.

[10] E. Curry, S. Scerri, T. Tuikka, Data Spaces: Design, Deployment, and Future Directions, Springer International Publishing, Cham, 2022, p. 1–17. doi:10.1007/978-3-030-98636-0_1.

[11] O. Corcho, M. Eriksson, K. Kurowski, M. Ojsteršek, C. Choirat, M. van de Sanden, F. Coppens, EOSC interoperability framework: Report from the EOSC Executive Board Working Groups FAIR and Architecture, Technical Report, EOSC Executive Board Working Groups FAIR and Architecture, 2021. doi:10.2777/620649.

[12] C. Jonquet, N. Grau, D4.2 - FAIR semantic artefact lifecycle from engineering, to sharing, 2025. doi:10.5281/zenodo.14643279.

[13] Pontus-X, Compute-to-Data with Nautilus, 2025. URL: https://docs.pontus-x.eu/docs/use-cases/compute.

[14] R. Garcıa, R. Gil, Agriculture data space for sovereign data sharing and semantic integration, in: Workshop on Semantic Interoperability in Data Spaces, Budapest, Hungary, 2024. URL: https://semantic.internationaldataspaces.org/wp-content/uploads/2024/08/Agriculture-Data-Space-for-Sovereign-Data-Sharing-and-Semantic-Integration.pdf.

[15] F. Gast, V. Berchtenbreiter, S. Dumss, R. Gehrer, P. Weißenbach, M. Weigold, M. Grafinger, Automatic publication of data to data and service ecosystems from the shopfloor, Procedia CIRP 130 (2024) 1589–1597. doi:10.1016/j.procir.2024.10.287.

[16] Ocean Protocol Foundation, Metadata: How can you enhance data discovery?, 2024. URL: https://docs.oceanprotocol.com/developers/metadata.

[17] V. L. Lemieux, R. Gil, F. Molosiwa, Q. Zhou, B. Li, R. Garcia, L. D. L. T. Cubillo, Z. Wang, Clio-X: AWeb3 Solution for Privacy-Preserving AI Access to Digital Archives, arXiv (2025). doi:10.48550/ARXIV.2507.08853.

[18] Data Spaces Support Centre, Data Models, 2025. URL: https://dssc.eu/space/BVE2/1071255252/Data+Models.

[19] SEMIC Support Centre, Application Profiles: What Are They and How to Model and Reuse Them Properly?, Interoperable Europe, 2024. URL: https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/application-profiles-what-are-they-and-how-model-and-reuse-them-properly-look-through-dcat-ap.

[20] N. Thalhath, M. Nagamori, T. Sakaguchi, Metadata application profile as a mechanism for semantic interoperability in FAIR and open data publishing, Data and Information Management 9 (2025) 100068. doi:10.1016/j.dim.2024.100068.

[21] W3C Dataset Exchange Working Group, Data Catalog Vocabulary (DCAT) – Version 3, W3C Recommendation, 2024. URL: https://www.w3.org/TR/vocab-dcat-3/.

[22] A. Morejón, L. de Espona, A. Berenguer, D. Tomás, J.-N. Mazón, Extension of Data Catalog Vocabulary for Federating Open Datasets in Data Spaces, in: Advances in Databases and Information Systems, volume 2676 of *Communications in Computer and Information Science*, Springer, 2026, pp. 108–117. doi:10.1007/978-3-032-05727-3_11.

[23] SEMIC Support Centre, DCAT-AP 3.0.0: DCAT Application Profile for Data Portals in Europe, Interoperable Europe Portal, 2024. URL: https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/release/300.

[24] A. Perego, V. Cetl, A. Friis-Christensen, M. Lutz, GeoDCAT-AP: Representing geographic metadata using the DCAT application profile for data portals in europe, in: Proceedings of the Joint UNECE/UN-GGIM Europe Workshop on Integrating Geospatial and Statistical Standards, Stockholm, Sweden, 2017. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC107410.

[25] E. C. J. R. Centre., ISO & GeoDCAT-AP metadata implementation Pilot: final report phase 1 (October 2024 – June 2025), Publications Office, LU, 2025. doi:10.2760/9289518.

[26] J. Masó, A. Brobia, R. Palma, I. Elicegui, Making Geospatial Data Assets Discoverable and Accessible in the Green Deal Data Space, AGILE: GIScience Series 6 (2025) 38. doi:10.5194/agile-giss-6-38-2025.

[27] C. Jonquet, J. Graybeal, S. Bouazzouni, M. Dorf, N. Fiore, X. Kechagioglou, T. Redmond, I. Rosati, A. Skrenchuk, J. L. Vendetti, M. Musen, members of the OntoPortal Alliance, Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology, volume 14266, Springer Nature Switzerland, Cham, 2023, p. 38–58. doi:10.1007/978-3-031-47243-5_3.

[28] C. Jonquet, N. Grau, M4.4 - Review of Semantic Artefact Catalogues and guidelines for serving FAIR semantic artefacts in EOSC, Technical Report, Zenodo, 2024. doi:10.5281/zenodo.12799796.

[29] C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé Yeumo, V. Emonet, J. Graybeal, M.-A. Laporte, M. A. Musen, V. Pesce, P. Larmande, AgroPortal: A vocabulary and ontology repository for agronomy, Computers and Electronics in Agriculture 144 (2018) 126–143. doi:10.1016/j.compag.2017.10.012.

[30] B. Kihal, C. Corre, C. Jonquet, D. Szabo, J. Roucou, S. Aubin, Leveraging AgroPortal Ontologies to Ease Metadata Completion and Data Discovery in Data INRAE, Technical Report, Zenodo, 2024. doi:10.5281/zenodo.14191078, FAIR-IMPACT Project.

[31] I. Subirats-Coll, K. Kolshus, A. Turbati, A. Stellato, E. Mietzsch, D. Martini, M. Zeng, AGROVOC: The linked data concept hub for food and agriculture, Computers and Electronics in Agriculture 196 (2022) 105965. doi:10.1016/j.compag.2020.105965.

[32] M. Hauff, L. M. Comet, P. Moosmann, C. Lange, I. Chrysakis, J. Theissen-Lipp, FAIRness in Dataspaces: The Role of Semantics for Data Management, in: Proceedings of the Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 3705 of *CEUR Workshop Proceedings*, 2024. URL: https://ceur-ws.org/Vol-3705/paper01.pdf.

[33] R. David, V. Alexiev, P. Ivanov, W. van den Berg, J. P. Wijbenga, M. Stornebrink, Federated Vocabulary Hubs as a Building Block for Semantic Layers in Data Spaces, in: Proceedings of the Third International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 4007 of *CEUR Workshop Proceedings*, 2025. URL: https://ceur-ws.org/Vol-4007/06short.pdf.

[34] M. Stäbler, P. Moosmann, P. Dittmer, D. Wang, F. Köster, C. Lange, Semantic Data Link: Bridging Domain-Specific Needs with Universal and Interoperable Semantic Models, in: Proceedings of the Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, volume 3705 of *CEUR Workshop Proceedings*, 2024. URL: https://ceur-ws.org/Vol-3705/short13.pdf.

[35] P. Derycke, B. J. Barros, N. M. Schutte, C.-A. Vande Catsyne, M. Bargeman Fonseca, Designing DCAT-AP Extensions for Common European Data Spaces: The EHDS HealthDCAT-AP Case Study, in: Proceedings of the NeXt-generation Data Governance Workshop 2025 (NXDG 2025), co-located with SEMANTiCS 2025, volume 4064 of *CEUR Workshop Proceedings*, Vienna, Austria, 2025. URL: https://ceur-ws.org/Vol-4064/NXDG25-paper5.pdf.

[36] M. Scrocca, L. Molinas Comet, B. Witsch, D. M. Mustafa, C. Lange, M. Comerio, P. Lubrich, mobilityDCAT-AP: A Metadata Specification for Enhanced Cross-Border Mobility Data Sharing, in: The Semantic Web: ESWC 2025 Satellite Events, volume 15719 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 226–243. doi:10.1007/978-3-031-94578-6_13.

[37] R. A. Deshmukh, D. M. Mustafa, G. Toubekis, B. T. Arnold, C. Gillmann, S. Decker, C. Lange, Semantic Information Modeling for Dataspaces: Culture IM, SIM Framework, and AP-first Methodology for Application Profile Development, Manuscript submitted to Semantic Web, 2026. URL: https://www.semantic-web-journal.net/system/files/swj4099.pdf, preprint.

[38] A. Tchechmedjiev, A. Abdaoui, V. Emonet, S. Zevio, C. Jonquet, SIFR annotator: ontology-based semantic annotation of French biomedical text and clinical notes, BMC Bioinformatics 19 (2018) 405. doi:10.1186/s12859-018-2429-2.

[39] N. H. Shah, N. Bhatia, C. Jonquet, D. Rubin, A. P. Chiang, M. A. Musen, Comparison of concept recognizers for building the open biomedical annotator, BMC Bioinformatics 10 (2009) S14. doi:10.1186/1471-2105-10-S9-S14.

[40] C. Jonquet, S. Bouazzouni, G. Alviset, N. Fiore, N. Karam, B. Kihal, C. Pierkot, M. Pulieri, I. Rosati, Federated FAIR Semantic Artefacts Discovery and Search with OntoPortal Federation, in: The Semantic Web – ISWC 2025 - 24th International Semantic Web Conference, Nara, Japan, November

2–6, 2025, Proceedings, Part II, volume 16141 of *Lecture Notes in Computer Science*, Springer, Nara, Japan, 2025, pp. 434–450. doi:10.1007/978-3-032-09530-5_25.

[41] European Commission, Directorate-General for Digital Services, SHACL validator, 2026. URL: https://www.itb.ec.europa.eu/shacl/any/upload, interoperability Test Bed service page. Accessed 2026-03-06.

[42] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, F. Ciravegna, Semantic annotation for knowledge management: Requirements and a survey of the state of the art, Journal of Web Semantics 4 (2006) 14–28. doi:10.1016/j.websem.2005.10.002.

[43] E. A. Papoutsoglou, D. Faria, D. Arend, E. Arnaud, I. N. Athanasiadis, I. Chaves, F. Coppens, G. Cornut, B. V. Costa, H. Ćwiek-Kupczyńska, B. Droesbeke, R. Finkers, K. Gruden, A. Junker, G. J. King, P. Krajewski, M. Lange, M.-A. Laporte, C. Michotey, M. Oppermann, R. Ostler, H. Poorter, R. Ramírez-Gonzalez, v. Ramšak, J. C. Reif, P. Rocca-Serra, S.-A. Sansone, U. Scholz, F. Tardieu, C. Uauy, B. Usadel, R. G. F. Visser, S. Weise, P. J. Kersey, C. M. Miguel, A.-F. Adam-Blondon, C. Pommier, Enabling reusability of plant phenomic datasets with MIAPPE 1.1, New Phytologist 227 (2020) 260–273. doi:10.1111/nph.16544.