DAOnt: A Formal Ontology for EU Data Act Compliance

Sheyla Leyva-Sánchez[1,*], Fabian Linde[1], Meem Arafat Manab[1], María Poveda-Villalón[1] and Víctor Rodríguez-Doncel[1]

[1]*Universidad Politécnica de Madrid, Madrid, Spain*

Abstract

The EU Data Act establishes comprehensive rules governing data access and sharing across business-to-consumer (B2C), business-to-business (B2B), and business-to-government (B2G) contexts. This paper presents a comprehensive ontology for the EU Data Act, enabling reasoning over data sharing agreements through machine-readable representations. The DAOnt ontology reuses elements from three established ontologies, LKIF-Core, ODRL, and DPV, to capture the normative structure of the Data Act.

The ontology captures the main concepts and relationships in the Regulation, and it also operationalises three articles to facilitate compliance checking: Article 4(1) (B2C user access rights), Article 8(6) (B2B trade secret exceptions) and Article 19(2)(a) (B2G competitive use prohibitions).

The ontology supports compliance checking through SPARQL queries that return obligations, permissions, and prohibitions, allowing organisations to verify whether data-sharing agreements meet the requirements of the EU Data Act and to assess conditions such as FRAND obligations. By representing key legal concepts in RDF, our work helps bridge the gap between the legal provisions of the Data Act and their computational interpretation. The complete ontology, along with example instances and queries, is available online.

Keywords

EU Data Act, Legal Ontology, Compliance Verification

1. Introduction

The EU Data Act[1] has become applicable on September 12, 2025, introducing a comprehensive regulatory framework for data access and sharing across the European Union. The regulation addresses three primary scenarios: business-to-consumer (B2C) data sharing where users gain access to data generated by their connected products (Chapter II, Articles 3-7), business-to-business (B2B) data sharing requiring user authorisation and fair, reasonable, and non-discriminatory (FRAND) terms (Chapter III, Articles 8-13), and business-to-government (B2G) data sharing under exceptional need circumstances (Chapter V, Articles 14-22).

Organisations face significant compliance challenges due to the complexity of the Regulation. The Data Act contains over 50 articles with intricate interdependencies: B2B sharing requires user authorisation per Article 8(4), must satisfy FRAND conditions per Articles 8-12, yet allows trade secret exceptions per Article 8(6). The manual compliance check is error-prone and does not scale, particularly for small and medium enterprises (SMEs) that lack dedicated legal resources. Data space architectures such as GAIA-X require interoperable contracts, while cloud providers need automated policy enforcement mechanisms.

The gap between legal text and computational enforcement creates practical barriers to compliance. Natural language provisions require human interpretation, leading to inconsistent implementations between organisations and jurisdictions. Without machine-readable representations, compliance checking remains a manual, costly process vulnerable to human error. Furthermore, existing compliance tools

CLAIRVOYANTS Workshop, Co-located with ESWC 2025

*Corresponding author.

sheyla.leyva.sanchez@upm.es (S. Leyva-Sánchez); fabian.linde@upm.es (F. Linde); meem.manab@upm.es (M. A. Manab); m.poveda@upm.es (M. Poveda-Villalón); vrodriguez@fi.upm.es (V. Rodríguez-Doncel)

0009-0007-9762-4045 (S. Leyva-Sánchez); 0009-0004-0856-8036 (F. Linde); 0000-0002-2336-4160 (M. A. Manab); 0000-0000-0000-0000 (M. Poveda-Villalón); 0000-0003-1076-2511 (V. Rodríguez-Doncel)

[1]Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act).

focus predominantly on GDPR and AI Act requirements, leaving the Data Act, despite its September 2025 applicability date, largely unaddressed by automated compliance solutions.

Machine-readable regulatory frameworks offer several advantages: automated reasoning can determine which obligations apply to specific scenarios, semantic queries can identify contract violations, and formal verification can validate policy compliance before deployment. Such frameworks enable affordable compliance tools for SMEs and support the interoperability requirements of European data spaces.

This paper addresses the research question: *Can compliance checking with the EU Data Act requirements be automated through formal ontology and semantic reasoning?*

Our contribution is threefold. First, we present a comprehensive ontology for the EU Data Act, which integrates three established standards (LKIF-Core, ODRL, and DPV) to formally represent regulatory requirements. Second, we demonstrate proof-of-concept formalizations for Articles 4(1), 8(6), and 19(2)(a), showing how different deontic modalities (mandatory obligations in B2C, permissive exceptions in B2B, and absolute prohibitions in B2G) can be captured and reasoned over. Third, we provide executable reasoning examples using SPARQL queries over RDF-formatted knowledge bases, demonstrating practical automated compliance checking capabilities.

Paper Structure

The remainder of the paper is structured as follows. Section 2 reviews related work on legal ontologies, data governance and data sharing ontologies, compliance approaches, and regulatory compliance ontologies, identifying the research gap that motivates our contribution. Section 3 presents the methodology, including the ontology engineering process and the integration of external vocabularies. Section 4 introduces the proof of concept, detailing the formalisation of Articles 4(1), 8(6), and 19(2)(a) and demonstrating automated reasoning examples in B2C, B2B and B2G compliance scenarios. Section 5 evaluates the ontology in terms of coverage, expressiveness, and interoperability. Finally, Section 6 concludes the article and Section 7 outlines the directions for future work.

2. Related Work

The development of an ontology for the EU Data Act builds upon two complementary research lines: the use of ontologies to represent legal and regulatory knowledge, and the formalisation of compliance mechanisms enabling automated or semi-automated assessment of normative requirements. The first line concerns the design of conceptual models that capture legal concepts such as rights, obligations, and permissions, providing a shared semantic foundation for interoperability and reasoning. These models are typically expressed using Semantic Web standards such as RDF and OWL, which ensure compatibility with existing ecosystems and support the representation of machine-readable normative structures. The second focusses on computational approaches to verify whether data processing or sharing activities comply with legal norms, using technologies such as SPARQL, SHACL, logic programming, or rule-based systems. This section reviews previous work in both areas, highlighting their relevance to the modelling of data-sharing obligations and justifying the design choices adopted in our ontology.

2.1. Legal ontologies

Early legal ontologies were designed primarily as conceptual frameworks to represent the structure of legal knowledge rather than as operational data models. Their main purpose was to facilitate understanding, interoperability, and knowledge sharing within the legal domain, rather than to support good metadata or even machine-executable compliance checking. This focus on conceptual modelling distinguished them from later ontologies, such as those used in data governance and compliance, which aim to be directly operationalised through technologies like OWL reasoners, SHACL, or SPARQL.

Comprehensive surveys by Casellas [1] and Rodrigues et al. [2] provide an overview of this evolution. Casellas' foundational work in 2011 describes the emergence of legal ontologies as efforts to formalise

legal concepts, case structures, and argumentation patterns, often within the contexts of knowledge management and document retrieval. In contrast, the more recent survey emphasises the shift toward lightweight, interoperable ontologies integrated with Semantic Web standards and designed to support automation, such as compliance checking and legal analytics.

Among the first and most influential initiatives, LKIF-Core (the Core Ontology of the Legal Knowledge Interchange Format) [3] stands out as a foundational ontology to represent legal reasoning and normative structures. Developed within the ESTRELLA project, LKIF-Core provides an upper-level schema for capturing concepts such as norms, acts, and roles, inspired by legal theory and deontic logic. Although not originally intended as an operational compliance model, LKIF-Core established a reusable conceptual foundation that has informed subsequent vocabularies. Its modular and extensible design continues to make it a reference point for aligning new legal ontologies with established conceptual categories.

LegalRuleML [4] complements ontology-based representations with a formal markup language to express legal norms, rules, and reasoning structures. While LKIF-Core provides a conceptual vocabulary, LegalRuleML offers a syntactic and logical framework to encode the prescriptive aspects of law–obligations, permissions, prohibitions, and exceptions–using an XML-based serialisation aligned with RuleML [5]. Its primary objective is to enable the interchange of machine-readable legal rules across systems, preserving their temporal and defeasible characteristics. However, unlike lightweight ontologies expressed in RDF, LegalRuleML was not conceived as a data model for operational use within Linked Data environments, but rather as a rule representation language supporting formal reasoning. For this reason, it bridges conceptual and computational layers, yet its adoption has remained limited outside specialised rule-based reasoning systems.

2.2. Ontologies for Data Governance and Data Sharing

The Open Digital Rights Language (ODRL) is a W3C specification, provided as two Recommendations [6][7], that provides a flexible and extensible policy expression language to represent permissions, prohibitions, and obligations associated with the use of digital assets. Originally conceived to manage usage rights in digital content distribution, ODRL has evolved into a general-purpose vocabulary to express policies that govern access, sharing, and usage of data and services. The activity in the domain of data spaces, closely related to our data act, is actually burgeoning. ODRL adopts an RDF-based model, enabling interoperability within the Semantic Web ecosystem and allowing policies to be linked to resources, parties, and actions. Although not specifically designed for legal compliance, the normative structure of ODRL – centred on the concepts of permission, prohibition, and duty – makes it particularly suitable for modelling data-sharing agreements and other legally relevant policies and has been widely reused and extended in data governance and privacy-related vocabularies.

DaPReCo (Data Protection Regulatory Compliance Ontology) [8] is a formal ontology designed to represent the legal concepts and obligations of the General Data Protection Regulation (GDPR), with a focus on enabling rule-based compliance assessment . Based on LegalRuleML and FOL-based formalizations, DaPReCo encodes the logical structure of GDPR provisions, allowing automated reasoning over obligations, rights, and legal bases for data processing. Beyond DaPReCo, several other ontologies have sought to capture the semantics of the GDPR at varying levels of abstraction and operationality.

The Data Privacy Vocabulary[2] (DPV), developed under the W3C Data Privacy Vocabularies and Controls Community Group, provides a lightweight RDF-based model for representing data processing purposes, legal bases and data subject rights, emphasising interoperability and practical deployment. Other initiatives-such as GConsent [9] or PrOnto [10] focus on narrower aspects such as consent modelling or personal data categories. Collectively, these ontologies illustrate a clear evolution from conceptual representations of privacy norms toward operational models that support machine-readable compliance verification.

[2] https://w3id.org/dpv/

2.3. Compliance approaches

Several conceptual approaches have been proposed to formalise and operationalise legal and regulatory compliance. DALICC (Data Access and Licence Interoperability for Content and Contracts) extends ODRL to model licencing and data-sharing policies in a way that supports interoperability across domains, bridging normative concepts with actionable constraints. In contrast, ontologies such as DaPReCo demonstrate how compliance checks can be performed by encoding GDPR rules into formal representations that allow reasoning engines to infer whether a given scenario satisfies obligations and permissions. Another widely used strategy involves SHACL (Shapes Constraint Language), which validates RDF data against predefined constraints, providing a declarative and relatively lightweight mechanism for ensuring compliance with structural or policy requirements. Some approaches go further by transforming ontology-based representations into other formal logic systems, such as first-order logic, Prolog, or Answer Set Programming (ASP), enabling advanced automated reasoning and defeasible rule evaluation. Finally, SPARQL offers a simpler yet effective means to query RDF representations directly, allowing organisations to verify specific obligations, prohibitions, or permissions without the overhead of full logic-based reasoning systems. Each of these approaches balances expressivity, computational complexity, and ease of integration, shaping the design choices for operational compliance ontologies.

2.4. Regulatory Compliance Ontologies

Automated compliance check approaches span multiple paradigms. Semantic logic-based approaches include First-Order Logic representations and frameworks such as I-SNACC that achieve 95.2% precision and 100% recall [11]. SHACL-based compliance checking shows increasing adoption for validating regulatory requirements. Policy language–based approaches leverage ODRL and LegalRuleML to formalise rights and obligations in a machine-interpretable way. An ODRL Compliance Profile has also been proposed to demonstrate effective coverage when combined with DPV [12]. Recent innovations include the use of Defeasible Logic Programming to handle contradictions and incompleteness, Answer Set Programming with event calculus formalisms, and hybrid approaches combining machine learning and rule-based reasoning.

AI Act ontologies emerged rapidly following the regulation's adoption, including AIRO (AI Risk Ontology, 2022[13]) providing OWL2 representation of AI system risks, VAIR (Vocabulary of AI Risks, 2023) for risk assessments, and TAIR (Trustworthy AI Requirements Ontology, 2024) modelling AI Act clauses and ISO standards.

2.5. Research Gap: The Missing Data Act Ontology

Despite the Data Act entering into force on January 11, 2024, with core provisions applying on September 12, 2025, no dedicated Data Act ontology currently exists. While DPV v2.0 includes Data Act support on its roadmap, this work remains incomplete. This represents an urgent gap for organisations that must comply without machine-readable specifications or automated compliance checking tools.

Current ontologies cannot express critical Data Act scenarios: connected product data rights, multi-tier data sharing chains (user → data holder → third party), conditional compensation mechanisms, FRAND term verification, switching rights for cloud services, or exceptional need provisions for public sector access. Our work addresses this gap through the first comprehensive formalisation of the Data Act.

3. Methodology

Traditional compliance verification relies on manual legal review or ad-hoc procedural code, both of which scale poorly as regulatory obligations become more numerous, interconnected, and dynamic. Ontologies provide a more robust alternative by formalising legal provisions into a machine-interpretable representation that reduces ambiguity and supports consistent, repeatable reasoning.

In our approach, the first stage consists of a comprehensive reading of the Data Act and internal discussions on how to formalise its key concepts, obligations, and legal actors. Based on this analysis, the regulation's requirements are modelled as explicit semantic structures with well-defined legal concepts and their normative relations. This semantic formalisation captures the regulatory logic in a form that software systems can reliably process. SPARQL queries then operationalise these structures, identifying compliance-relevant patterns with minimal procedural code.

This methodology transforms compliance assessment from a subjective, manual task into an automated and auditable process, improving transparency, scalability, and maintainability. As regulations evolve, ontology-based representations also offer a flexible foundation that facilitates continuous monitoring without extensive re-engineering.

3.1. Ontology Engineering Process

The development process follows the NeOn methodology [14], an iterative scenario-based framework well suited to modelling complex legal domains. We conducted a systematic requirement elicitation process through three complementary activities. First, a close reading and detailed annotation of the EU Data Act (Regulation 2023/2854) identified the main actors, core concepts, and their relationships throughout the regulation. Second, we formulate competency questions and organise them into two categories: general CQs ensuring core conceptual coverage (e.g. "Who shares data with whom, and under which agreement?"), and article-specific CQs supporting the proof-of-concept formalisation (e.g., "Which data holders violate Article 4(1) by failing to provide the user-requested data from connected products?"). The complete catalogue is provided in Table 1. The full Ontology Requirements Specification Document (ORSD) is included in Annex A. Third, we performed an interoperability analysis to maximise reuse of established vocabularies—importing foundational legal concepts from LKIF-Core and extending data governance notions using DPV and ODRL—thereby avoiding redundant or conflicting definitions.

Tool-supported refinement was carried out throughout the iterative modelling cycle. Chowlk [15] was used to automatically transform conceptual diagrams into formal OWL representations, and selected patterns from the Ontology Design Patterns catalogue ensured reusable and well-structured modelling decisions. Validation and refinement were performed through a consistency check with OWL reasoners (HermiT and ELK), complemented by SPARQL queries that verified that all competency questions were answerable.

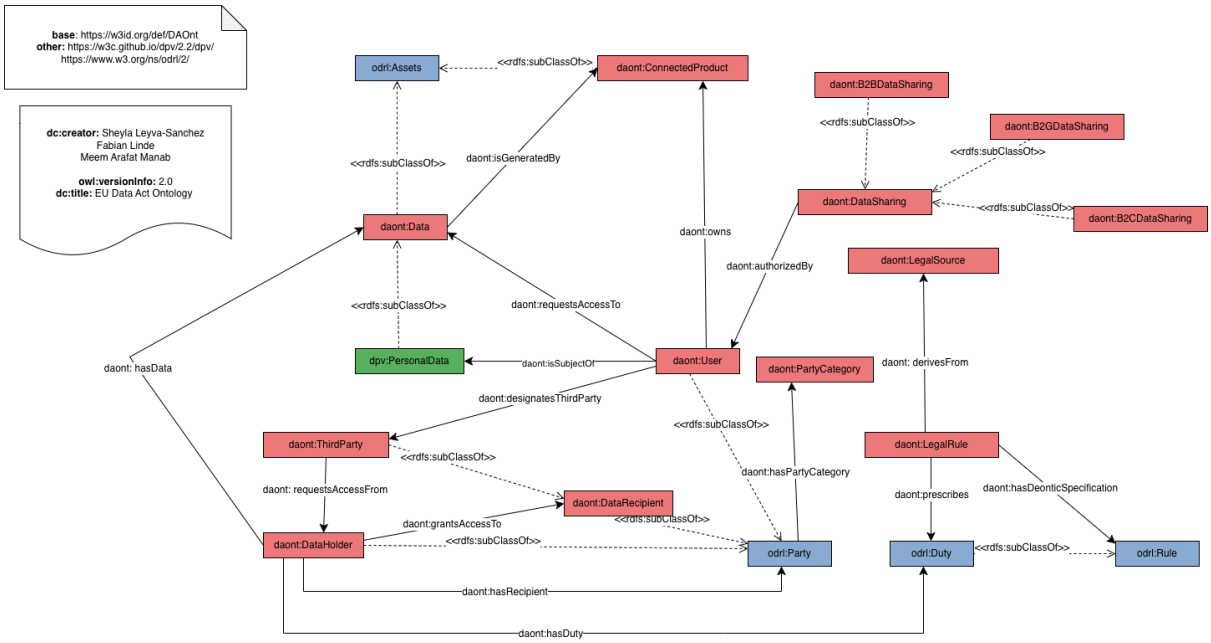

Figure 1: A top-level diagram of the ontology, depicting most relevant classes. (Red tiles signify classes from the *DAOnt* ontology, blue from *ODRL*, green from *DPV*)

For documentation, we used Widoco [16], which generates human-readable specifications, metadata, and publication-ready HTML documentation. Finally, the ontology was published following the best practices of W3C using OnToology [17], guaranteeing persistent hosting, versioning, and automated quality evaluation.[3][4]

CQG1. General CQs (9 CQs)	CQG2. CQs Related to PoC Articles (3 CQs)
CQ1. Who shares data with whom, and under what agreement? CQ2. What actions has a party performed? CQ3. What obligations does a party have? CQ4. Who manufactured a product? CQ5. Who uses a product? CQ6. What service does a product provide? CQ7. When does a legal rule apply? CQ8. Where does data come from? CQ9. Who holds the data?	CQ10. Which data holders have violated Article 4(1) by failing to provide requested data from connected products? CQ11. Which data holders have violated Article 8(6) by refusing data sharing without providing trade secret justification? CQ12. Which public sector bodies have violated Article 19(2)(a) by developing competing products or services using data obtained through B2G data sharing?

Table 1
Competency questions grouped by general coverage (CQG1) and article-specific requirements for the proof of concept (CQG2).

4. Proof of Concept: Three Regulatory Articles

In this section, we present a proof of concept that illustrates how DAOnt enables automated compliance checking across different regulatory contexts of the Data Act. We analyse three representative articles (4(1)[5], 8(6) [6], 19(2)(a) [7]), each corresponding to a distinct compliance modality—mandatory obligations, conditional permissions, and absolute prohibitions—and spanning the B2C, B2B, and B2G settings of the regulation. These examples demonstrate how the ontology captures normative structures with sufficient expressiveness to support automated reasoning. Table 2 provides a comparative overview of the formal representations used in each case.

4.1. Automated Reasoning Examples Across B2C, B2B, and B2G

To demonstrate compliance verification capabilities, we developed six contract examples instantiating both compliant and non-compliant scenarios across the three data sharing contexts: B2C, B2B, and B2G. The class instances representing these six scenarios are available in the GitHub repository[8].

B2C Context (Code Snippet 1). In the violation scenario, `charlie` (a `ConsumerUser`) owns `smartWatch1`, which generates `charlieHealthData`. Charlie requests access to these data, but `watchManufacturer` (a `DataHolder` and `Manufacturer`) fails to perform any `DataProvision`

[3]Public URI: https://w3id.org/def/daont
[4]Development repository: https://github.com/oeg-upm/DAOnt
[5]Data holders shall make readily available data, as well as the relevant metadata necessary to interpret and use those data, accessible to the user without undue delay, of the same quality as is available to the data holder, easily, securely, free of charge, in a comprehensive, structured, commonly used and machine-readable format and, where relevant and technically feasible, continuously and in real-time
[6]An obligation to make data available to a data recipient shall not oblige the disclosure of trade secrets.
[7]A public sector body, the Commission, the European Central Bank, a Union body or a third party receiving data under this Chapter shall not use the data or insights about the economic situation, assets and production or operation methods of the data holder to develop or enhance a connected product or related service that competes with the connected product or related service of the data holder
[8]https://github.com/oeg-upm/DAOnt/tree/main/compliance-checks/contracts

Table 2
Comparison of three Data Act articles

Aspect	Article 4(1) B2C	Article 8(6) B2B	Article 19(2)(a) B2G
Legal Text	"shall make data available without undue delay"	"not required to make data available where data constitute trade secrets"	"shall not use data to develop competing products"
Deontic Type	Obligation (`odrl:Duty`)	Permission (`odrl:Permission`)	Prohibition (`odrl:Prohibition`)
Condition	User requests data access	B2B request + trade secret	None (unconditional)
Action	*provideDataAction*	*refuseDataAccessAction*	*UseDataToDevelop-CompetingProduct*
Exception	None	*containsTradeSecret-Condition*	None
Assignee	Data holder (must act)	Data holder (may refuse)	Public sector body (cannot act)

action, thereby violating the mandatory obligation in Article 4(1). Data sharing is governed by `contract_charlie`.

To detect this violation automatically, the SPARQL query in Listing 1 searches for B2C data-sharing cases where a `ConsumerUser` requests access to data linked to a product they own or use, and checks whether the associated `DataHolder` performs the required `DataProvision` action. The `FILTER NOT EXISTS` block encodes the notion of a *missing obligation*: if no `DataProvision` action is found, the pattern is flagged as non-compliant. The corresponding compliant scenario satisfies the obligation, and therefore the query does not return violations.

Code Snippet 1: B2C: Missing Obligation

```
SELECT ?holder ?data
WHERE {
  ?sharing a da:B2CDataSharing;
        da:governedBy ?c .
  ?c dpv:hasRecipient ?user .
  ?user a da:ConsumerUser;
      da:ownsOrUses ?product;
      da:requestsAccessTo ?data.
  ?holder a da:DataHolder;
       dpv:hasData ?data .
  FILTER NOT EXISTS {
    ?holder da:performsLegalAction ?provision .
    ?provision a da:DataProvision .
  }
}
```

B2B Context (Code Snippet 2). In the violation scenario, `factoryOwnerAcme` (an `EnterpriseUser`) owns `industrialRobot1`, which generates `robotData1`. The factory owner authorises `autoRepair` (an `AftermarketServiceProvider` and `DataRecipient`) to access these data through `agreement247`, governed by `contract247`, which specifies FRAND-compliant terms (`frand247` with `isFair`, `isReasonable`, and `isNonDiscriminatory` all set to true). Despite these conditions, `robotManufacturer` (a `DataHolder`) refuses to share the data without providing a trade secret justification, violating the conditional sharing requirement in Article 8(6). The compliant scenario provides access under the same FRAND conditions and includes a valid justification whenever disclosure is refused.

To detect this violation automatically, the SPARQL query in Listing 2 searches for B2B data-sharing cases where a `DataHolder` does not perform the required `DataProvision` action and simultaneously fails to provide a `containsTradeSecret` justification. The first `FILTER NOT EXISTS` block encodes

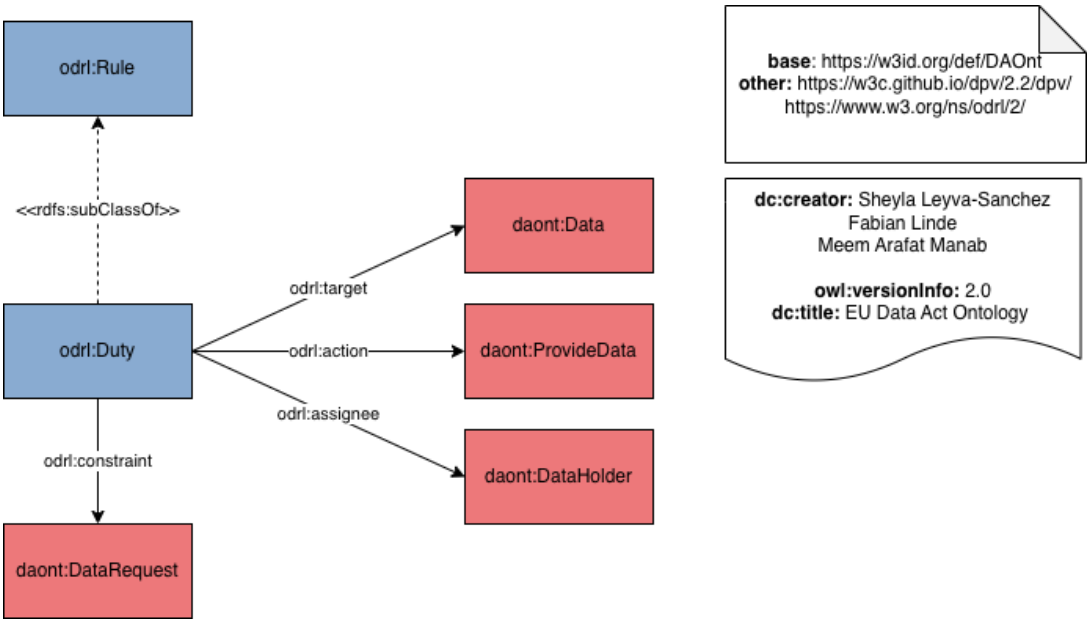

Figure 2: Diagram depicting the principal classes and properties involved in DA Art. 4(1), based on the ODRL design pattern. (Red tiles signify classes from the *DAOnt* ontology, blue from *ODRL*)

the absence of the mandatory action, while the second captures the absence of the legally permitted exception. When both conditions hold, the query precisely identifies the non-compliant behaviour described in Article 8(6).

Code Snippet 2: B2B: Refusal Without Trade Secret Justification

```
SELECT ?holder ?recipient
WHERE {
  ?sharing a da:B2BDataSharing;
         da:governedBy ?c;
         da:authorizedBy ?user .
  ?user da:ownsOrUses ?product .
  ?product da:generatesData ?data .
  ?holder a da:DataHolder;
        dpv:hasData ?data .
  ?c dpv:hasRecipient ?recipient .
  FILTER NOT EXISTS {
    ?holder da:performsLegalAction ?provision .
    ?provision a da:DataProvision .
  }
  FILTER NOT EXISTS {
    ?data da:containsTradeSecret ?s .
  }
}
```

B2G Context (Code Snippet 3). In the violation scenario, `gonzalo` (a `ConsumerUser`) owns `healthMonitor1`, which generates `gonzaloHealthData`. The `healthAuthority` (a `PublicSectorBody` and `DataRecipient`) legitimately requests access to these data from the `healthDeviceManufacturer` under `publicHealthEmergency2024`, an instance of `ExceptionalNeed` authorising B2G sharing for pandemic monitoring. This exchange is governed by `contract191`. However, after obtaining the data, the authority performs `competitiveProductDevelopment1` (a `UseDataToDevelopCompetingProduct` action), thus using the data to develop a competing health application—an explicit violation of the absolute prohibition in Article 19(2)(a). The compliant scenario demonstrates lawful use restricted strictly to the declared public interest purpose.

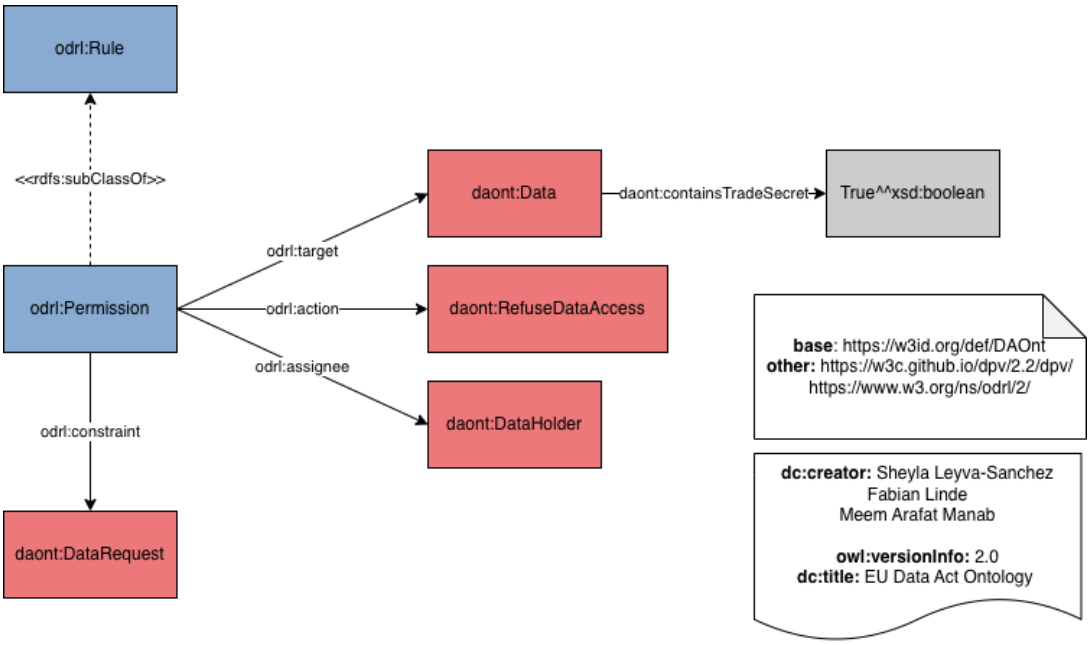

Figure 3: Diagram depicting the principal classes and properties involved in DA Art. 8(6), based on the ODRL design pattern. (Red tiles signify classes from the *DAOnt* ontology, blue from *ODRL*, gray signifies XML Schema Definitions)

To detect this violation automatically, the SPARQL query in Listing 3 searches for B2G data-sharing cases where a `PublicSectorBody` performs an action classified as `UseDataToDevelopCompeting-Product`. Unlike the B2C and B2B contexts, which require detecting *missing obligations* or *missing exceptions*, Article 19(2)(a) defines an absolute prohibition. For this reason, the query simply checks for the presence of the prohibited action; any match directly corresponds to non-compliance according to the regulation.

Code Snippet 3: B2G: Prohibited Action

```
SELECT ?publicBody ?action
WHERE {
  ?sharing a da:B2GDataSharing;
        da:governedBy ?c .
  ?c dpv:hasRecipient ?publicBody .
  ?publicBody a da:PublicSectorBody .
  ?holder a da:DataHolder;
       dpv:hasData ?data;
       dpv:hasRecipient ?publicBody .
  ?publicBody da:performsAction ?action .
  ?action a da:UseDataToDevelopCompetingProduct .
}
```

These queries illustrate how SPARQL operationalises the semantic structures defined in DAOnt to support automated compliance checking. By encoding mandatory obligations, conditional permissions, and absolute prohibitions as graph patterns, queries leverage the formal semantics of the ontology to detect violations directly from the data. This shows the practical value of combining ontology-driven modelling with query-based reasoning: compliance verification becomes scalable, transparent, and auditable, and regulatory requirements can be enforced without bespoke procedural logic.

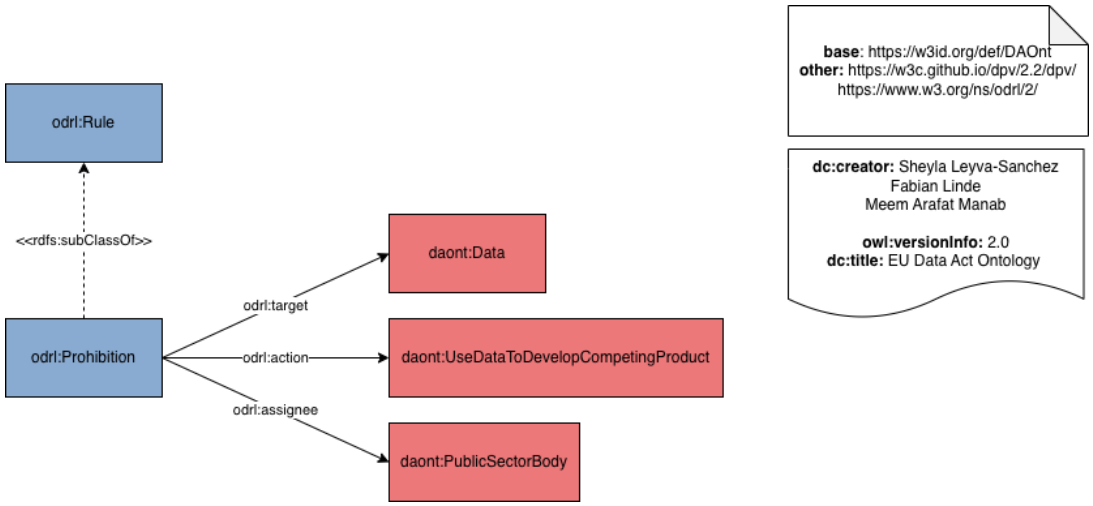

Figure 4: Diagram depicting the principal classes and properties involved in DA Art. 19(2)(a), based on the ODRL design pattern. (Red tiles signify classes from the *DAOnt* ontology, blue from *ODRL*)

4.2. Automated Compliance Verification Engine

These compliance verification queries are designed to detect Data Act violations through automated reasoning over RDF knowledge graphs. Such compliance checking may be needed in various contexts where data sharing occurs under regulatory obligations—for example, in data space infrastructures like Gaia-X[9], sectoral platforms for health or mobility, or enterprise data sharing APIs. In these environments, a compliance module monitors activities by creating RDF instances for relevant events (contracts, data requests, data provisions) representing actual cases with real parties, datasets, and timestamps. Instance creation is hybrid: systems automatically capture events from APIs and smart contracts, while humans or LLMs judge concepts that cannot be formalised symbolically—like whether a delay is "undue" (Article 4) or if products "compete" (Article 19(2)(a)). SPARQL queries are then executed periodically or events are triggered to detect violations. Critically, this relies on the Closed World Assumption (CWA): `FILTER NOT EXISTS` interprets missing instances as violations, treating absence as non-compliance—contrasting with OWL's Open World Assumption where missing information means unknown. Since Data Act violations occur through inaction, the CWA reasoning is essential for compliance auditing.

To demonstrate this compliance verification approach, we developed a first implementation prototype in Python, using the RDFLib library to manage and query the knowledge graph. This approach enables the efficient loading of the foundational DAOnt schema along with the synthetic contract instances into a unified RDF graph. By executing the SPARQL queries for compliance (as detailed in the preceding section) directly against this graph, we were able to automatically detect instances of non-compliance. To validate the system's efficiency, we implemented a prototype compliance dashboard using Streamlit. This tool transforms the output of semantic reasoning into user-friendly, auditable reports that demonstrate several key advantages as follows:

- Efficiency: The application processes all six synthetic contracts in milliseconds, confirming the high throughput of the ontology-driven approach. This high processing speed contrasts sharply with the time and manual effort a human auditor would require to review contracts against regulatory text.
- Interpretable Results: Results are inherently transparent due to their foundation in SPARQL graph pattern matching on the DAOnt. The system provides a direct, traceable explanation by identifying the specific SPARQL rule and the exact entities involved in the violation.

[9]In Gaia-X, these are "Compliance Services" within the Trust Framework.

- Auditable Oversight: The prototype facilitates immediate, interactive feedback for regulatory oversight by linking factual contractual data directly to the formal legal requirements encoded in DAOnt.

The operational prototype can be accessed at https://daont-verify.streamlit.app/

5. Evaluation and Discussion

5.1. Coverage Assessment

We evaluated ontology coverage against the Data Act's core chapters using OnToology's automated documentation pipeline. The current implementation covers 18 articles in three primary contexts: B2C user access rights (Chapter II, Articles 3-7), B2B data sharing with FRAND conditions (Chapter III, Articles 8-13), and B2G exceptional need provisions (Chapter V, Articles 14-22). Articles 4(1), 8(6), and 19(2)(a) serve as representative formalizations demonstrating the modalities of obligation, permission, and prohibition, respectively. The remaining gaps include cloud switching mechanisms (Chapter VI, Articles 23-31), interoperability requirements (Chapter VII, Articles 32-34), and smart contract provisions (Chapter VIII, Article 30).

5.2. Expressiveness Analysis

OOPS! (OntOlogy Pitfall Scanner) detected no critical errors in the ontology structure. Minor pitfalls include missing domain/range declarations for 7 properties (P11 warning) and absence of inverse relationships for 4 object properties (P13 warning), both intentional design decisions to maintain compatibility with DPV's lightweight RDFS approach rather than full OWL2-DL constraints. The ontology successfully expresses complex deontic scenarios including conditional permissions (trade secret exceptions requiring `containsTradeSecretCondition`), multi-party chains (user $\rightarrow$ data holder $\rightarrow$ third party recipient), temporal constraints (Art. 4(1) "without undue delay"), and compensation mechanisms (Art. 12 FRAND terms).

5.3. Interoperability Benefits, limitations and challenges

Integration with four established standards enables multiple deployment scenarios. DPV alignment allows direct reuse in existing GDPR compliance systems (e.g., consent management platforms already using DPV vocabularies). ODRL compatibility supports data space implementations like GAIA-X requiring machine-readable policies. SPARQL queries execute on standard triple stores (GraphDB, Apache Jena, Virtuoso) without custom reasoning engines. LKIF-Core alignment facilitates future integration with legal reasoning systems that model Hohfeldian relationships. OnToology-generated documentation (HTML, diagrams, metadata) reduces adoption barriers through comprehensive reference materials.

Three primary limitations constrain current applicability. First, semantic ambiguity in legal text requires human interpretation for edge cases—Article 8(6) "trade secrets" lacks precise boundaries, requiring domain experts to instantiate `containsTradeSecret` predicates. Second, dynamic policy evolution remains unaddressed; contract modifications require manual knowledge base updates rather than automated policy versioning. Third, enforcement mechanisms lie outside ontology scope—while SPARQL queries detect violations, actual enforcement (notifications, sanctions, dispute resolution) requires external system integration. Validation with legal practitioners and industry stakeholders remains ongoing, with preliminary feedback indicating usability challenges for non-technical users requiring simplified tooling layers.

6. Conclusions

This work introduces the first formal ontology designed specifically to support compliance with the EU Data Act, addressing a critical gap in the landscape of machine-readable regulatory frameworks. By reusing and aligning concepts from ODRL, DPV, and LKIF-Core, DAOnt provides a coherent semantic model to represent rights, obligations, permissions, and prohibitions in data sharing agreements in B2C, B2B, and B2G contexts.

The proof-of-concept scenarios demonstrate that semantic modelling combined with SPARQL reasoning enables practical, fine-grained compliance checking. Executable queries successfully detect missing obligations, unjustified refusals, and prohibited actions, illustrating how legal requirements can be operationalised directly on top of ontology-driven representations. This confirms the viability of ontologies as a foundation for transparent, auditable, and scalable compliance verification.

Beyond immediate use cases, DAOnt lays the groundwork for future tools that support affordable compliance automation, interoperable contractual templates for data spaces, and policy-aware data governance infrastructures. As organisations prepare for the enforcement of the Data Act, ontology-based approaches such as the one presented here offer a promising pathway toward robust, trustworthy, and machine-processable regulatory compliance.

7. Future Work

Several directions remain open to expand the ontology and its compliance verification capabilities. First, we plan to conduct structured consultations with legal experts to refine the modelling of rights, obligations, permissions, and exceptions, ensuring that the ontology captures the nuances of legal interpretation and domain-specific practice. Second, the current proof-of-concept focusses on three representative articles; future work will expand the compliance checker to cover a broader range of provisions from the Data Act, including cross-cutting obligations, sector-specific rules, and complex exceptional-need scenarios. Third, we aim to explore the semi-automatic generation of Data Act–compliant contractual templates driven by our ontology, enabling machine-generated contracts to serve as executable test cases and reusable reference examples for automated compliance verification.

Acknowledgments

This research is funded by the European Union's Horizon 2020 programme under the Marie Skłodowska-Curie grant agreement No. 101169409 (HARNESS) and project PID2024-159504OB-I00, MICIU/AEI /10.13039/501100011033 and FEDER, UE.

Declaration on Generative AI

Artificial intelligence was used exclusively for spelling checks and grammar improvement in the preparation of this manuscript.

References

[1] N. Casellas, Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge, Law, Governance and Technology Series, Springer Netherlands, 2011. URL: https://books.google.es/books?id=JBR2zq-voVQC.

[2] C. M. de Oliveira Rodrigues, F. L. G. de Freitas, E. F. S. Barreiros, R. R. de Azevedo, A. de Almeida Filho, Legal ontologies over time: A systematic mapping study, Expert Systems with Applications 130 (2019) 12–30. URL: https://www.sciencedirect.com/science/article/pii/S0957417419302398. doi:https://doi.org/10.1016/j.eswa.2019.04.009.

[3] R. Hoekstra, J. Breuker, M. D. Bello, A. Boer, Lkif core: Principled ontology development for the legal domain, in: J. Breuker, P. Casanovas, M. C. A. Klein, E. Francesconi (Eds.), Law, Ontologies and the Semantic Web, volume 188 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2009, pp. 21–52. doi:10.3233/978-1-58603-942-4-21.

[4] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, A. Paschke, Legalruleml: Xml-based rules and norms, in: Rule-Based Modeling and Computing on the Semantic Web, volume 7018 of *Lecture Notes in Computer Science*, Springer, 2011, pp. 298–312. URL: https://link.springer.com/chapter/10.1007/978-3-642-24908-2_30. doi:10.1007/978-3-642-24908-2_30.

[5] N. Bassiliades, G. Governatori, A. Paschke (Eds.), Rule Representation, Interchange and Reasoning on the Web, volume 5321 of *Lecture Notes in Computer Science*, Springer, 2008.

[6] R. Iannella, S. Villata, ODRL Information Model 2.2, W3C Recommendation REC-odrl-model-20180215, World Wide Web Consortium (W3C), 2018. URL: https://www.w3.org/TR/2018/REC-odrl-model-20180215/, editors: Renato Iannella (Monegraph), Serena Villata (INRIA).

[7] R. Iannella, M. Steidl, S. Myles, V. R. Doncel, ODRL Vocabulary & Expression 2.2, W3C Recommendation REC-odrl-vocab-20180215, World Wide Web Consortium (W3C), 2018. URL: https://www.w3.org/TR/2018/REC-odrl-vocab-20180215/, editors: Renato Iannella (Monegraph), Michael Steidl (IPTC), Stuart Myles (AP), Víctor Rodríguez Doncel (UPM).

[8] L. Robaldo, C. Bartolini, G. Lenzini, The dapreco knowledge base: Representing the gdpr in legalruleml, in: Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020), European Language Resources Association, Marseille, France, 2020, pp. 5688–5697. URL: https://aclanthology.org/2020.lrec-1.698/.

[9] H. J. Pandit, C. Debruyne, D. O'Sullivan, D. Lewis, GConsent: A consent ontology based on the GDPR, in: Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019), volume 11503 of *Lecture Notes in Computer Science*, 2019, pp. 270–282. URL: https://doi.org/10.1007/978-3-030-21348-0_18. doi:10.1007/978-3-030-21348-0_18.

[10] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, L. Robaldo, Pronto: Privacy ontology for legal reasoning, in: A. Kő, E. Francesconi (Eds.), Electronic Government and the Information Systems Perspective (EGOVIS 2018), volume 11032 of *Lecture Notes in Computer Science*, Springer, 2018, pp. 139–152. URL: https://doi.org/10.1007/978-3-319-98349-3_11. doi:10.1007/978-3-319-98349-3_11.

[11] J. Wu, X. Xue, J. Zhang, Invariant signature, logic reasoning, and semantic natural language processing (nlp)-based automated building code compliance checking (i-snacc) framework, Journal of Information Technology in Construction 28 (2023).

[12] M. De Vos, S. Kirrane, J. Padget, K. Satoh, Odrl policy modelling and compliance checking, in: International Joint Conference on Rules and Reasoning, Springer, 2019, pp. 36–51.

[13] D. Golpayegani, H. J. Pandit, D. Lewis, Airo: An ontology for representing ai risks based on the proposed eu ai act and iso risk management standards, in: Towards a knowledge-aware AI, IOS Press, 2022, pp. 51–65.

[14] M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, A. Gangemi, The NeOn Methodology for Ontology Engineering, Springer, 2012.

[15] S. Chávez-Feria, C. A. Iglesias, Chowlk: From uml to owl automatically, in: Proceedings of the 21st International Semantic Web Conference (ISWC 2022) – Demo Track, 2022.

[16] D. Garijo, Widoco: A wizard for documenting ontologies, in: The Semantic Web–ISWC 2017, Springer, 2017, pp. 94–102.

[17] D. Chaves-Fraga, D. Garijo, M. Poveda-Villalón, O. Corcho, Ontoology: A tool for automating ontology documentation, evaluation, and publication, in: Proceedings of the 6th Workshop on Linked Data on the Web (LDOW 2013), 2013.

A. Full Ontology Requirements Specification Document (ORSD)

<table>
<tr><th colspan="2">Ontology Requirements Specification Document</th></tr>
<tr><td colspan="2">1. Purpose</td></tr>
<tr><td colspan="2">The purpose of this machine-readable model of the EU Data Act is to enable the verification of compliance of data sharing agreements against Data Act requirements.</td></tr>
<tr><td colspan="2">2. Scope</td></tr>
<tr><td colspan="2">The ontology models the main regulatory provisions of the Data Act related to B2C, B2B, and B2G data sharing, focusing on deontic relations such as duties, permissions, and prohibitions.</td></tr>
<tr><td colspan="2">3. Implementation Language (optional)</td></tr>
<tr><td colspan="2">Web Ontology Language (OWL)</td></tr>
<tr><td colspan="2">4. Intended End-Users (optional)</td></tr>
<tr><td colspan="2">User 1. Regulatory compliance officers seeking to verify contractual conformity.
User 2. Data space and cloud service providers implementing machine-readable policies.
User 3. Legal experts and policymakers analyzing normative implications.</td></tr>
<tr><td colspan="2">5. Intended Uses</td></tr>
<tr><td colspan="2">Use 1. Compliance verification for data sharing agreements.
Use 2. Automated detection of violations and noncompliance.
Use 3. Semantic querying of legal rights and duties.</td></tr>
<tr><td colspan="2">6. Ontology Requirements</td></tr>
<tr><td colspan="2">a. Non-Functional Requirements</td></tr>
<tr><td colspan="2">NFR 1. The ontology must be based on the international, European or de-facto standards in existence or under development.</td></tr>
<tr><td colspan="2">b. Functional Requirements: : Lists or tables of requirements written as Competency Questions and sentences</td></tr>
<tr><td>CQG1. General CQs (9 CQs)</td><td>CQG2. CQs Related to PoC article (3 CQs)</td></tr>
<tr><td>CQ1. Who shares data with whom, and under what agreement?
CQ2. What actions has a party performed?
CQ3. What obligations does a party have? CQ4. Who manufactured a product?
CQ5. Who uses a product?
CQ6. What service does a product provide?
CQ7. When does a legal rule apply?
CQ8. Where does data come from
CQ9. Who holds the data</td><td>CQ10. Which data holders have violated Article 4(1) by failing to provide requested data from connected products?
CQ11. Which data holders have violated Article 8(6) by refusing data sharing without providing trade secret justification?
CQ12. Which public sector bodies have violated Article 19(2)(a) by developing competing products or services using data obtained through B2G data sharing?</td></tr>
</table>

Figure 5: Complete Ontology Requirements Specification Document (ORSD) for DAOnt.