NGI ONTOCHAIN

Blockchain for the Next Generation Internet

ONTOROPA

ONTOROPA
D1. 'STATE OF THE ART AND AMBITION'

23/04/2021

Grant Agreement No.: 957338
Call: H2020-ICT-2020-1

Topic: ICT-54-2020
Type of action: RIA

ONTOROPA

D1. 'STATE OF THE ART AND AMBITION'

DUE DATE	23/04/2021
SUBMISSION DATE	23/04/2021
TEAM	ONTOROPA
VERSION	2.0
AUTHORS	M.Mercedes Martínez-González, Pompeu Casanovas, María-Luisa Alvite-Díez, Núria Casellas
COLLABORATORS	Inma Gutiérrez, María Ruíz, David Sanz, Amador Aparicio, Sergio Miguel

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

EXECUTIVE SUMMARY

OntoROPA is a project proposal to facilitate smart privacy legal compliance using technology capable of providing semantics, intelligence, and trust. This is an innovative proposal for the targeted marketplace—i.e. the legal compliance marketplace— since there are no solutions capable of simultaneously providing the three properties mentioned. OntoROPA deals with the creation and maintenance of a critical piece of legal compliance required by the GDPR, the Records of Processing Activities (ROPA).

OntoROPA pursues genuine regulatory compliance, with full meaning and legal validity. To this end, the ability to certify registrations, ROPAs, and to validate them are fundamental elements. In short, to prove that the ROPAs are true and that they meet all requirements that the GDPR requires. This is also generating legal evidence.

OntoROPA ambition is to innovate in legal compliance checking and monitoring, bootstrapping blockchain technology to show that it can also be used for privacy compliance in the new LawTech market. Innovation in legal compliance will be achieved by providing legal value to digital artifacts and procedures created to comply with legal data protection requirements at regional, national and European level. This is something that current tools in the legal compliance market do not provide. Blockchain technology has been put under question by privacy experts, drafters, and rulers because its distributed nature is not entirely compatible with GDPR requirements. OntoROPA will provide to blockchain technology the way to address the issues that have been raised and to remove the technical and legal obstacles. In addition, doing so, it will create a specific market niche, generating a secure and trustworthy legal ecosystem with economic value.

OntoROPA proposes the creation of a knowledge graph, a RDF graph, to handle information about ROPAs. It combines building a professional ontology that will be part of this graph with the collection and management of the specific knowledge of the community of privacy and data protection experts—mainly including lawyers, legal advisors and scholars, data protection officers, and rulers who are proficient in the creation and manipulation of ROPAs. This will trigger the future implementation of other procedures capable of inferring new knowledge from the available one. Certification and trusthworthiness are included as requirements in the very first phases of its design. Blockchain based solutions will be included and headed towards this aim.

OntoROPA can benefit from the ONTOCHAIN projects providing innovative and usable methodologies for the following purposes : (i) Certification of the genesis and origins of a ROPA; (ii) validation; (iii) proof of proactivity; (iv) identity management; (v) management of semantics; (vi) and smart community support.

Synergies with other ONTOCHAIN projects are detailed in section 4. There may be more projects with which such a collaboration is possible. The criteria we followed to elaborate this section is to highlight the projects with whom we started planning an effective cooperation. In section 5 we present the conclusions achieved at this time slot of project development.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

LIST OF FIGURES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

LIST OF TABLES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

ABBREVIATIONS

GDPR	General Data Protection Regulation
ROPA	Records Of (Personal) Data Processing Activities

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

1 INTRODUCTION

Providing Records of Personal Data Processing Activities (ROPAs) is currently a mandate of the General Data Protection Regulation (GDPR) Regulation (EU) 2016/679) (EU 201657), which rules ROPAs in Article 30. ROPAs should not be independent and isolated pieces of information. They should be reliable sources of information, linked, available for intelligent knowledge extraction. Our proposal aims to provide metadata about ROPAs that can be assessed with automatic and intelligent processes. The specific objectives of this proposal are: 1) To obtain a standard ontology for ROPAs; 2) Being able to share trustworthy and open information about ROPAs, ready to be exploited by intelligent processes, in a community-based ecosystem. Our thesis is that linking high-quality data about ROPAs will allow for intelligent extraction of knowledge from these data protection items of information, flexible comparisons of ROPAs, and intelligent processes that assist the inspection of ROPAs.

Our solution has two main components. The first one is a domain ontology, a ROPA ontology that collects the knowledge about the information ROPAs should contain. Our ambition is that this ontology gets a standard when the ONTOCHAIN project ends. The second component is the technical framework that will support the organization of a community that shares distributed, reliable, connected ROPAs, the ontological ROPAs, on top of which intelligent processes can be applied. Elements of this ecosystem are: 1) the information architecture; 2) the infrastructure that provides technical measures for linking, sharing, trust, and validation; 3) the community that participates as users and creators of knowledge and information and will be key for the standardization of the OntoROPA ontology. The ontology and ROPAs will be open, available as RDF and OWL files. Linked Data provides the ability to share and link them, and blockchain the technology to provide trust.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2 STATE OF THE ART

2.1 Regulatory and legal compliance

In a broad definition, compliance is the conformance of human or artificial behaviour with a set of rules, norms, principles, or values. In the Internet of Things (IoT), compliance has also been bootstrapped, because if humans must be compliant, so must be cyber-physical systems (CPS), autonomous and intelligent systems (AI/S), socio-technical systems (STS), and socio-cognitive technical systems (SCTS). Regulatory and legal compliance should be carefully distinguished. *Regulatory compliance* refers to the concept, languages and methodologies developed within the business, commercial and corporate fields to design, control and monitor in advance business processes and activities. *Legal compliance* refers to the formal developments that can be deemed 'legal' according to the norms, principles, and jurisdictions of regional, national, international, and transnational legal systems. They certainly converge, but the meanings of the two notions should be kept separate, as some requirements must be added for legal compliance be accorded from official bodies.

This is linked to the *Compliance by Design* (CbD) schemes that have been developed in the corporate business field since the beginning of the century to cope with the constraints set by the Sarbanes-Oxley Act (2002), a US Federal law that laid down new requirements for public company boards and accounting firms. There is some confusion ins this regard. In computer science literature *regulatory compliance* also denotes "the act and process on ensuring adherence to laws" that involves "discovering, extracting and representing different requirements from laws and regulations that affect a business process." (Akhigbe et al., 2015).

In the past twenty years, several formal languages have been developed to carry out these tasks, following a variety of methodologies and techniques described many times in the literature on the subject, within four main fields: (i) deontic logic (temporal deontic logic and computational tree logic), (ii) Petri nets, (iii) graph-based business modelling—BPMN, event-driven process chain (EPC diagrams), unified modelling language (UM)—, (iii) goal-oriented languages, (iv) and languages for the semantic web (legalXML, legalRuleML).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Figure 1 maps these different trends.

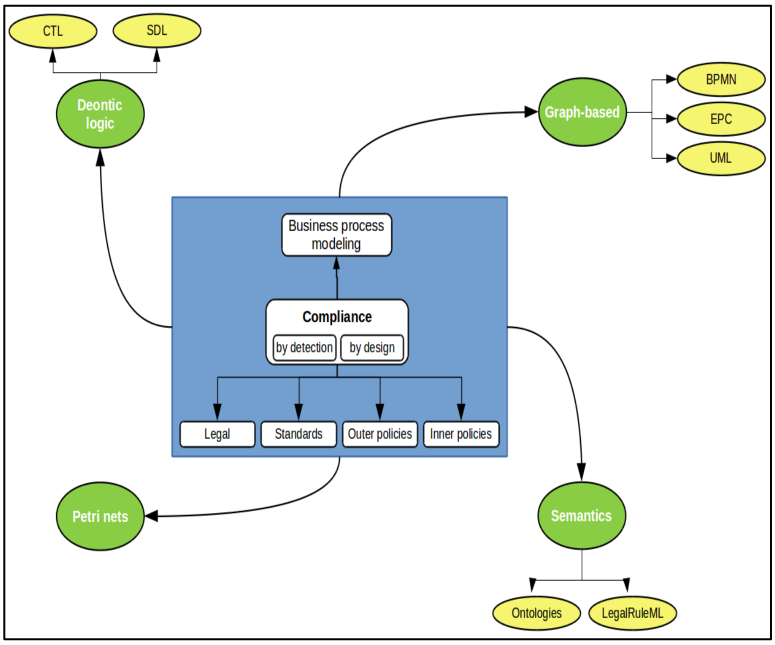


Figure 1. Languages for business compliance models.
SDL: Standard Deontic Logic, CTL: Computer Tree Logic, BPMN: Business Process Model and Notation, EPC: Event-driven Process Chain, UML: Unified Modeling Language

Legal compliance represents an extension of these epistemic approaches outside of the business and corporate areas to encompass all fields of regulation under the laws—private, commercial, corporate, industrial, administrative, criminal, public etc. I.e. basically embracing all substantive and formal rights that are implemented through the rule of law. As said, this is adding complexity to the whole compliance process. Thus, we have suggested elsewhere (Casanovas et al., 2017)to differentiate: (i) (Automated) regulatory compliance and (semi-automated) legal compliance, (ii) Compliance by Design (CbD) and Compliance through Design (CtD). The latter are focused on *legal knowledge*, defining some more requirements based on the properties of normative legal systems (hierarchy, consistency, effectivity, etc.) to encompass the social and institutional dimensions of regulations within the IoT —from documentary legal interpretation to the coordination of all stakeholders and the relation between citizens and the law. According to our results, years 2009 (in the middle of the last financial

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

crisis, Fig. 2) and 2020-21 (because of the enforced implementation of GDPR) are the tipping points of the growing interest of industry and researchers to find CbD and CtD solutions (Akhigbe et al., 2019).

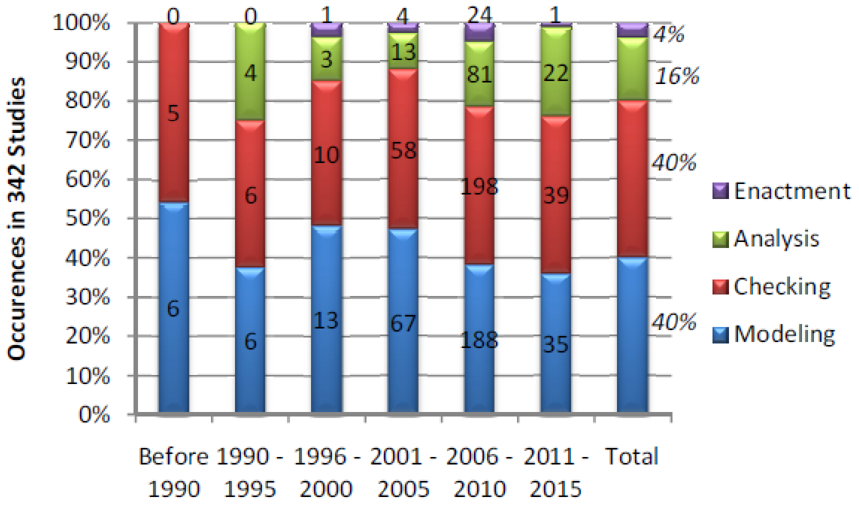


Figure 2. State of the regulatory compliance tasks over time (until 2015). Source: Akhibe, et al.(2015).

Hashmi, Governatori, Lam, and Wynn (2018) have identified the next challenges. Without being exhaustive: (i) the expressivity of formal languages to represent normative contents; (ii) the extraction of formal rules expressed in natural language, (iii) coping with multi-jurisdictional requirements, (iv) how to deal with control flow-structure, (v) integrating rules with processes, (vi) handling violations, (vii) dealing with model evolution, (viii) handling the performance and complexity of the models, (ix) and their usability, understandability, and explainability.

The last feature, 'explainability'—or *explicability*, assembling explanatory means and accountability (Floridi & Cowls, 2019)—is important here, because it deals with ethical principles, and ethical principles deal with Artificial Intelligence. Transparency and explainability will be even more important in the future, according to the first draft of the next *EU Artificial Intelligence Regulation.*[1]

[1] Recital n. 7 of the first Draft read: "https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence, including on ethics, liability, copyright, artificial intelligence in criminal matters, and artificial intelligence in education, culture and

This applies as well to blockchain solutions. Many recent proposals to make compatible GDPR provisions and blockchain solutions do exist (Campanile et al., 2021; Faber wt al., 2019). Some of them explicitly focus on blockchain, RuleML and the law of contracts, trying to model the legal notion of contract in a way that could be acceptable by a judge (Governatori et al. 2018).

An example of the intersection of legal requirements and blockchain technology are the recent cases of copyright infringement derived from the creation and sale of non-fungible tokens (NFTs) by unverified accounts in platforms such as Twinci.

NTFs have been viewed and promoted as a tool to empower artists and ensure revenue for their creations. The platform itself describes the sale of digital art as NTFs in the following manner:

> *As an artist, by tokenizing your work you both ensure that it is unique and brand it as your work. The actual ownership is blockchain-managed.*

However, recent developments suggest that compliance with copyright laws and ensuring authorship/ownership of the digital art contained in the NFT was not an integral part of the NTF technological solution being implemented.

In a way, NFTs represent a meta-ownership concept, which relies on code to allow for ownership-like digital distribution, exhaustion, remunerated resale, and enforcement within the context of a blockchain-based system. In doing so, NFTs offer an appealing new remuneration model for creators. However, for the most part, the affordances of NFTs are not accompanied by matching legal effects as far as copyright law is concerned. This creates significant challenges for creators, other

the audio-visual sector. The European Parliament resolution on a framework of ethical aspects of artificial intelligence, robotics and related technologies specifically recommends to the Commission to propose a legislative action to harness the opportunities and benefits of artificial intelligence, but also to ensure protection of ethical principles." The European Parliament and the Council of the European Union. *Regulation on a European Approach For Artificial Intelligence*. Recitals 5 and 6 of the final first version of the draft, officially released on April 21st, flesh out this same idea, adding market and legal constraint. R 5: "[…] rules regulating the placing on the market and putting into service of certain AI systems should be laid down, thus ensuring the smooth functioning of the internal market and allowing those systems to benefit from the principle of free movement of goods and services. By laying down those rules, this Regulation supports the objective of the Union of being a global leader in the development of secure, trustworthy and ethical artificial intelligence, as stated by the European Council33, and it ensures the protection of ethical principles, as specifically requested by the European Parliament". We deem R 6 truly relevant: "*The notion of AI system should be clearly defined to ensure legal certainty, while providing the flexibility to accommodate future technological developments*". Data Governance and Data Protection are specially focused in several articles (e.g. Art. 10).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

rights holders, and users if their expectation is that an NFT transaction on a blockchain will mirror an off-chain transaction for an equivalent work.[2]

Several artists denounced that NFTs created from their digital art (including deceased artists) were being sold on the platform without their knowledge and permission.

The platform has recently issued a statement indicating that there are unverified users in their platform and advising other users not to purchase from them. They have also announced that "KYC solutions will be deployed to improve transparency, non-tampering, and authenticity".

2.2 Semantic interoperability

2.2.1 Methodologies for ontology design

A methodology may be understood as an organized set of procedures and guidelines for, in this case, aiding and guiding the development of an ontology during its life-cycle or some parts of it. An ontology engineering methodology considers, management (e.g., scheduling), development (e.g., conceptualization and formalization), and support activities (e.g., knowledge acquisition and evaluation). These iterative activities conform a methodological life-cycle to support the construction of ontologies, from their specification to their implementation and maintenance.

Current ontology methodologies are mostly adapted from existing software and knowledge system engineering methodologies, and offer guidance towards knowledge acquisition, ontology development (design and conceptualization), formalization, evaluation, evolution and maintenance. For an extensive research on ontology methodologies see Casellas(2011).

2.2.2 Main Ontology Methodologies

Most methodologies generally follow a cyclic modelling approach, an iterative and incremental cycle of steps. The preparatory step includes

[2] http://copyrightblog.kluweriplaw.com/2021/04/14/the-rise-of-non-fungible-tokens-nfts-and-the-role-of-copyright-law-part-i/; http://copyrightblog.kluweriplaw.com/2021/04/22/the-rise-of-non-fungible-tokens-nfts-and-the-role-of-copyright-law-part-ii/. Enlaces a twitter: https://twitter.com/twinciio/status/1383862381441257476?s=20; https://twitter.com/loishh/status/1383820509381464068?s=20

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

activities such as the initial feasibility study and the specification of requirements. The development step comprehends knowledge acquisition activities, together with ontology conceptualization, formalization, evaluation and refinement, and the application step, which includes ontology implementation and maintenance activities. See **Appendix A**, Table 1.

2.2.3 Expert Knowledge as Methodology

Although most ontology methodologies have been highly influenced by the existing standards and methodologies regarding software and systems design, few of the revised methodologies have been deeply influenced by the standards and methods set towards a **human-centred perspective to systems (ontology) design**, domain expert-centred design.

Most ontology methodologies may involve domain experts and users at some stages of the development process (mainly knowledge acquisition and evaluation), although none of the above-mentioned methodologies describes a **complete expert-centred perspective towards ontology engineering**.

In general, no reference is made towards ensuring that the knowledge modelled in the ontology is, in fact, **shared amongst the experts or professionals of the domain**.

Human-centred software design and user validation are highly standardised processes which include participation in and evaluation of the general development of software, systems and products, the analysis of their usability, the documentation provided and the quality of their use.

In this project, we take into account the detailed modelling guidelines from Noy and McGuinness (Ontology Development 101) and Methontology, but include expert-centred and empirically-oriented methods towards professional legal knowledge acquisition, and usability (shareability) evaluation towards the construction of the ROPA Ontology.

The methodological steps will follow the general cyclic iterative and incremental approach: specification of requirements, knowledge acquisition, conceptualization, formalization, evaluation and refinement.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

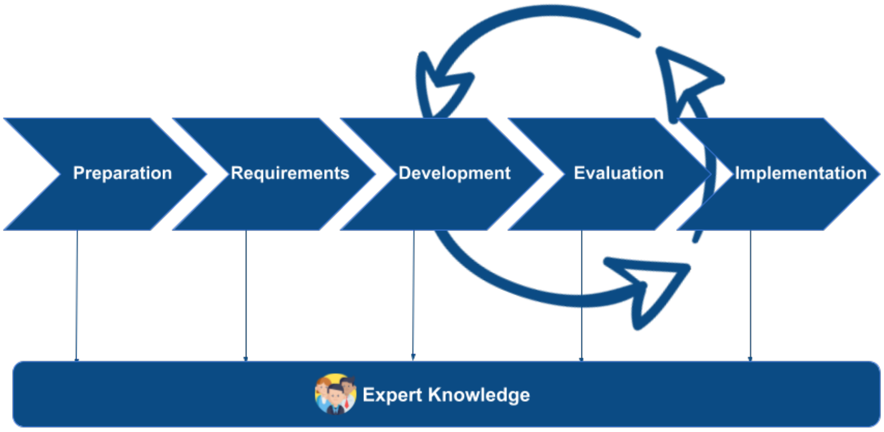


Figure 3. Methodological steps

2.2.4 Review of GDPR Ontologies

Since enactment of data protection regulations in the European Union and elsewhere, from the repealed *Data Protection Directive (DPD, Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*[3]) to the current *General Data Protection Regulation (GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC*[4]), many have pursued the encoding of their semantics for the development of smart data privacy compliant applications.

We mainly describe a selection of GDPR-related ontologies that are relevant to our project that are available for review and reuse, we also focus on their use of expert knowledge during their development. See the details in **Appendix B**. For extensive accounts of data protection related ontologies see Pandit(2020)and Esteves & Rodríguez-Doncel(2021).

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046
[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2.2.5 GDPR Expert Professional Knowledge

While there are many semantic models that focus on GDPR concepts, there are currently no ontologies that model GDPR expert professional knowledge with a focus on the ROPA maintenance and management required by data controllers and supervisors of the records.

Legend:

- ✕ No expert involvement | Not publicly available
- ☒ Limited expert involvement
- ✓ Extensive expert involvement | Publicly available

Table 1. GDPR ontologies

	Purpose	**Expert Knowledge**	**Formality| Expressivity**	**Publicly available**
SPECIAL usage policy language	**"This document specifies the SPECIAL usage policy language, which can be used to express both the data subjects' consent and the data usage policies of data controllers in formal terms, understandable by a computer, so as to automatically verify that the usage of personal data complies with data subjects' consent".**	✕	OWL2	✓
Data Privacy Vocabulary	**"The DPV is a vocabulary (terms) and an ontology (relationships) serialised using semantic-web standards to represent concepts associated with privacy and data protection, primarily derived from GDPR. It enables representation of which personal data categories are undergoing a what kind of processing by a specific data controller and/or transferred to some recipient for a particular purpose, based on a specific legal basis (e.g., consent, or other legal grounds such as legitimate interest, etc.), with specified technical and organisational measures and restrictions (e.g., storage locations and storage durations) in place".**	✕	RDF/OWL	✓
Policy Log Vocabulary	**"This document specifies splog, a vocabulary to log data processing and sharing events that should comply with a given consent provided by a data subject. We also model the consent actions related to consent giving and revocation."**	✕	RDF/OWL	✓

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

	Purpose	Expert Knowledge	Formality| Expressivity	Publicly available
DVP-GDPR	"he Data Privacy Vocabulary (DPV) provides terms (classes and properties) to describe and represent information related to processing of personal data. This extension extends the DPV and provides concepts specific to the obligations and requirements of the General Data Protection Regulation (GDPR). More specifically, it provides a taxonomy of legal bases and rights as defined within the GDPR."	✕	RDF/OWL	✓
GDPRov	"GDPRov (pronounced GDPR-Prov) is a linked data ontology for expressing provenance of consent and data lifecycles with a view towards documenting compliance".	✕	OWL2	✓
GConsent	"The ontology is based on an analysis of modelling metadata requirements related to the consent lifecycle for GDPR compliance. It allows modelling and representation of information related to compliance in an extensible and comprehensive manner."	✕	OWL2	✓
GDPRtEXT	"The GDPRtEXT ontology aims to provide a way to refer and use concepts defined by the General Data Protection Regulation (GDPR)."	✕	RDF/OWL(SKOS)	✓
Data Protection Ontology	"...the ontology will constitute the knowledge base from which the concepts to annotate the workflow model are extracted. Such an approach can provide benefits for a number of stakeholders: – data controllers would have a clearer view of their duties with respect to data protection in the context of their business; – the auditors would have a first-look model to assess the GDPR compliance; – DPAs would have a structured approach to detect potential violations."	☒	OWL	✓
PrOnto (Privacy Ontology)	"The PrOnto (Privacy Ontology) provides concepts regarding legal privacy compliance associated with data types and documents, agents and roles, processing purposes, legal bases, processing	☒	OWL	✕

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

	Purpose	Expert Knowledge	Formality| Expressivity	Publicly available
	operations, and deontic operations for modelling rights and duties."			
Compliance Ontology/Information Model Ontology/Policy Model Ontology	"The Compliance Ontology documented in detail in the Deliverable D3.1 "Compliance ontology specification" is a generic and, at the same time, highly expressive model ultimately grounded on the analysis of the GDPR that could be easily mapped to the underlying domain-specific information model of any organisation; it actually provides a high-level codification of the GDPR, by extracting the concepts that need to be addressed by the BPR4GDPR policy framework, as well as by the privacy-aware process reengineering."	✕	OWL	✕
Fiesta-Priv ontology	"a. Inspired by the GDPR requirements, we propose an IoT ontology built using available standards that enhances privacy, enables semantic interoperability between IoT deployments and supports the development of privacypreserving experimental IoT applications."	✕	OWL	✓
BIoT	"...the ontology provided insight into security properties to monitor vulnerabilities in the IoT ecosystem and blockchain network structure, thereby ensuring data integrity, confidentiality, and privacy. "	✓	OWL	✕

Table: Overview of most relevant ontology models for GDPR representation

The ROPA Ontology will provide an expert-driven semantic model for the automated management of ROPA.

2.3 Blockchain, privacy, and data protection

What does it mean to **participate in a blockchain transaction**?

The latest EU comparative *Report on EU Blockchain Ecosystems Developments* reads: "France is at the forefront of crypto asset recognition in Europe, having passed a legal framework for Initial Coin Offerings as early as 2016, followed by further legislative initiatives in 2017 and 2018. There is a relatively large number of blockchain companies in the country,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

with one of the world’s most successful hardware wallet providers (Ledger) headquartered in Paris.” (EU Report, 2020).
This experience is most valuable to structure our OntoRopa proposal on the effective implementation of data protection.

According to the French Commission Nationale de l'Informatique et des Libertés (CNIL) participants in blockchain technologies can be considered also as data controllers, because (i) they define the purposes (objectives pursued by the processing) and the means (data format, use of blockchain technology, etc.) of the processing, (ii) and hey have the right to write on the chain to decide to send data for validation by the miners. More specifically, the CNIL considers that the participant is a data controller:

- when the participant is a natural person, and the personal data processing operation is related to a professional or commercial activity (i.e. when the activity is not strictly personal);
- when the participant is a legal person (a company, bank, store, corporation, administration….) that registers personal data in a blockchain.(CNIL, 2018).

Permissionless blockchains are distributed, decentralised peer-to-peer networks in which everyone can participate interacting with unknown counterparties, trusted or not. The clear allocation of responsibilities that is required by GDPR are not present in this situation, as assessed by Michèle Fink’s study for the European Parliament on blockchain and data protection (Fink, 2019).

Some issues that have been already discussed are: (i) the assumption that in relation to each personal data point there is at least one natural or legal person; (ii) the uncertain contours of the notion of (joint)-controllership under the regulation; (iii) the assumption that data can be modified or erased where necessary, (iv) the question whether personal data that has been encrypted or hashed still qualifies as personal data, (v) the difficulty in determining whether data can be sufficiently anonymised, (vi) the GDPR requirement that personal data that is processed be kept to a minimum and only processed for purposes that have been specified in advance, (vii) blockchain’s impossibility to cope with the right to erasure, the 'right to be forgotten', as it has been deliberately designed to render the modification of data difficult or impossible; (ix) ambiguities in GDPR concepts such as “joint controller” or even “anonymisation” do not help ledger technologies to comply with the requirements of the Regulation. This entails that whoever intents to get into a blockchain transaction must have a proactive data protection attitude. Hence, the CNIL recommends identifying beforehand the data controller to comply with the “joint controller” of article 26

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957338

of he GDPR. (Otherwise, all participants could be considered data controllers).[5]

Fink's study recommends the adoption of regulatory guidance by EU experts, closing agreements between regulators and the private sector, and the elaboration of codes of conduct and certification mechanisms for blockchain technologies that should be "compliant by design". It is worth mentioning that these are "better regulations" mechanisms to be enhanced for harmonising both separate visions—centralised vs. decentralised governance principles.

See **Appendix C**, the Table that summarises the legal risks at stake, as set by the CNIL analysis of the situation.

By way of a precaution, to be on the safe side, the CNIL is crystal clear about the risks of not being fully compliant with GDPR provisions:

> "The CNIL recognizes the value of these solutions but, at this point, questions their ability to ensure a full compliance with the GDPR. This subject is one of the issues for which a reflection at the European level is essential." (CNIL, 2018, p.5)

Thus, the CNIL also recommends (i) "*to store any data in cleartext outside of the blockchain (such as, for example, on the data controller's information system)*"; (ii) "to store on the blockchain only a *proof of existence* of the data, in order of preference: (a) commitment, (b) hash generated from a keyed hash function, (c) cyphertext of the data; (iii) if none of these solutions can be implemented, "when justified by the purpose of the processing, and *when a DPIA has proven that the residual risks are acceptable*", data can be stored (a) either using a hash function without a key, or (b) in the absence of any other possibilities, in cleartext.

Likewise, with respect to the suitable measures to safeguard the data subject's rights [e.g. the rights of rectification and restriction] and "freedoms and legitimate interests", (i) "*the data subject should be able to obtain human intervention*, to express his or her point of view and to contest the decision after the smart contract has been performed", (ii) and "*the data controller should therefore provide the possibility*

[5] "Data subjects (i.e. those whose personal data is recorded on the blockchain) must know which entity they can refer to in order to effectively exercise their rights, and data protection authorities must have a contact point who can be held accountable for the processing carried out.", (CNIL, 2018, p.2).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

of human intervention allowing the data subject to contest the decision even if the contract has already been performed, and regardless of what is registered on the blockchain". This requires taking data subject's rights into account while writing the programme, i.e. prior to the implementation of a smart contract.

It is our contention that the recommendations provided by the CNIL encourage the usage of *homomorphic encryption*— an advanced method of encryption that enables the computation of cyphertexts—and *Zero-knowledge proofs*, i.e. solutions that can be used to provide a binary true/false answer without providing access to the underlying data.[6]

These recommendations are (partially) coincident as well with the experts' opinions gathered by the focus groups conducted by the British Blockchain Association (Schwerin, 2018). We reproduced its findings in Table 2.

Table 2. JBBA: A summary of hypotheses and compared rated Delphi results

H1: Blockchains have an impact on personal data.	• Electronic identity for which consumers create a separate identity for every digital service they are using, to which they can grant granular access rights for specific services (interoperability). • Blockchains help to improve documentation of personal data processes.
H2: Data protection regulations will have an impact on blockchains related to personal data.	• There should be minimum standards for security and the ability for users to manage consent. • Particular care towards personal data should be considered when dealing with digital avatars.
H3: Personal data cannot be stored on the blockchain directly, but indirectly.	• With regards to blockchains personal data is considered personally identifiable content, metadata and transactions.

[6] "Where zero-knowledge proofs are used, the blockchain indeed only shows that a transaction has happened, not which public key (as sender) transferred what amount to the recipient. It has moreover been pointed ought that zero knowledge proofs and homomorphic encryption have the potential to solve the conflict between data minimisation and the verifiability of data between many parties". (Fink, 2019, p.33).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

H4: Blockchains can be designed in a privacy-friendly manner by using the approach of privacy by design.	• In terms of privacy by design blockchains could be compliant but should not do it alone. • Basic design principles need to be established by open standards to ensure that blockchains maintain personal data integrity.
H5: Blockchains can help to solve (privacy) challenges accompanying the implementation of the new GDPR.	• All blockchain designers should be conscious of human rights, data protection and privacy as well as the need to consider how technology generally can protect the privacy of the individual without impeding technological progress.

These expert findings are similar but a bit more optimistic than CNIL's recommendations. What does it mean 'compliance' and specifically 'Compliance by Design' (CbD) in this context? To avoid misinterpretations, we should focus now on this issue.

2.4 OntoROPA methods for knowledge acquisition

2.4.1 Expert-based knowledge acquisition

As we pointed out in section *2.1 Methods for Ontology Design*, trends related to methodologies about knowledge acquisition in the formalisation of ontologies focused on data protection/privacy, specifically on the GDPR, are generally based on the modelling of the legal text (Data Privacy Vocabulary -DPV, 2021; Kirrane et al., 2018; Kost et al., 2011, 2012; Palmirani & Governatori, 2018; Palmirani et al., 2018). However, this project emphasises the use of expert knowledge as methodological starting point.

The knowledge acquisition stage is mainly based on the elicitation of expert knowledge, an approach also observed in the field of legal ontologies (Casellas, 2011; Sartor et al., 2011; Poblet et al. 2009, Rodrigues et al., 2019). We consider that this approach is innovative in this domain and much better suited to an application domain such as ROPAs and the qualitative research methods are considered appropriate to the type of knowledge the project aims to collect (Akhavan et al. 2018; Creswell, 2014; Dehghani et al., 2017).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2.4.2 Expert knowledge elicitation stages

For the development of the OntoROPA ontology we have established five different stages and methods for eliciting knowledge from experts:

1. ***Delimitation of the target group of experts***: the project will work with a target group of data protection officers and/or persons in charge of the ROPAs in Spain.
2. ***Survey***: A questionnaire with open-ended questions addressed to experts, staff responsible of ROPAs was developed to gather initial data.

   It was launched on the 12th of April and it will be open until the 3th of May. Accessible at https://forms.office.com/r/frLei8GtQH. The aim is to know the current state of know how related to the management of ROPAs, aspects such as data models, formats, technologies used, procedures, legitimizing bases of the GDPR, costs, etc.

   The preliminary results show heterogeneity in the procedures carried out, poor formalisation of content and lack of semantic interoperability, among others.

   - The procedures used for the drafting of ROPA processing operations are mostly manual.
   - Concerning the method of collecting or publishing information, PDF or HTML documents are the preferred.
   - The survey highlights the large amount of time required for the drafting and maintenance of the ROPA (an average of more than 10 weeks in drafting and annual maintenance time of more than 15 days).
   - 100% of respondents agree that it would be useful to have automated tools that would allow a more agile management, with standards that would simplify their work.
   - Automated management and the use of standards would mean cost savings for organisations.
3. ***Focus group***: this technique will be used to explore the opinions, knowledge, perceptions, details of processes, and concerns of experts involved in the project in regard to ROPAs. This interview method will facilitate the construction of knowledge in a collaborative way.

4. ***Competency questions:*** this step will specify what knowledge has to be entailed in the ontology as a set of requirements on the content as well as a way of scoping and delimiting the subject domain that has to be represented in OntoROPA ontology.

5. ***Evaluation:*** experts will be asked to validate the ontology at different stages of development.

2.4.3 Additional sources of data

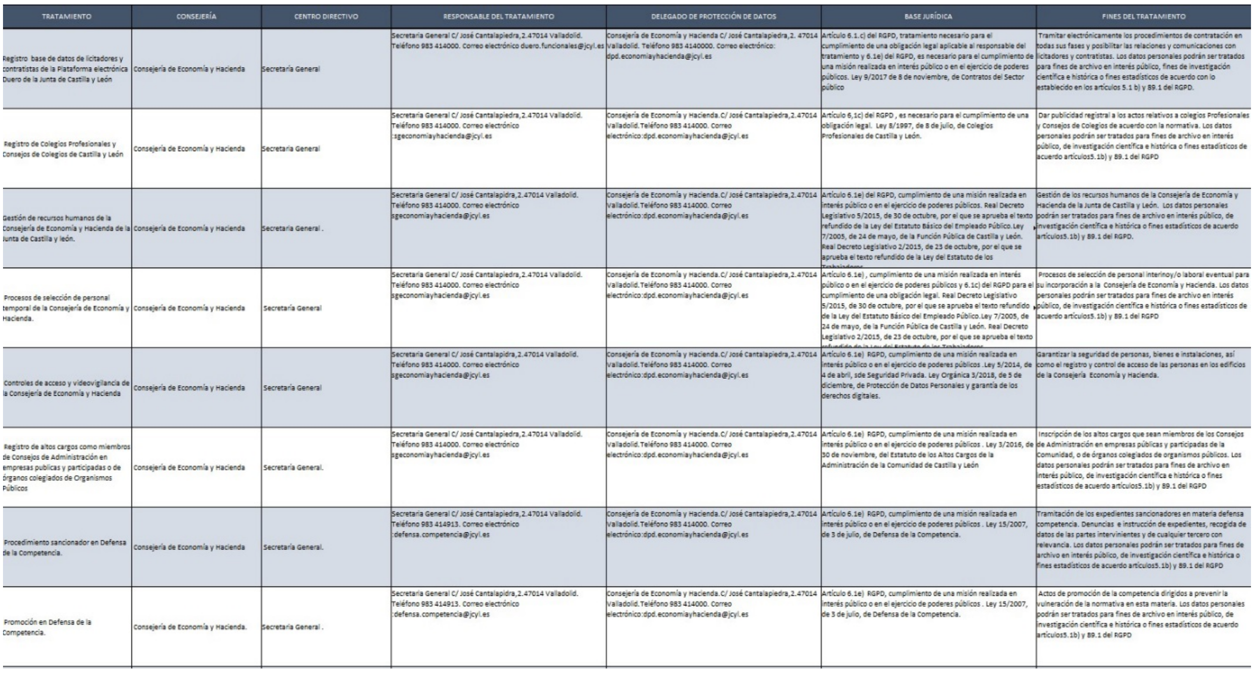
During knowledge acquisition we will also review other important data sources such as: published and available ROPA for direct analysis, the text of GDPR relevant to the obligations related to the records of processing activities, and the review and evaluation of related ontologies for reuse.

2.4.4 Published ROPA

Initial contacts with public administration officials, and information security experts indicate that the obligation of the creation of ROPA, Records of Processing Activities, is a task that many public administrations in Spain are not yet complying with. The same is true for private businesses and entities acting on the IoT.

The project has gathered an available list of published ROPA for analysis, in most cases they consist of spreadsheets with the information required by GDPR but there are also PDF files. These ROPAs are published on websites that do not have to be certified in any way.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338



TRATAMIENTO	CONSEJERÍA	CENTRO DIRECTIVO	RESPONSABLE DEL TRATAMIENTO	DELEGADO DE PROTECCIÓN DE DATOS	BASE JURÍDICA	FINES DEL TRATAMIENTO
Registro base de datos de licitadores y contratistas de la Plataforma electrónica Duero de la Junta de Castilla y León	Consejería de Economía y Hacienda	Secretaría General	Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico duero.funcionales@jcyl.es	Consejería de Economía y Hacienda C/ José Cantalapiedra,2. 47014 Valladolid. Teléfono 983 4140000. Correo electrónico: dpd.economiayhacienda@jcyl.es	Artículo 6.1.c) del RGPD, tratamiento necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento y 6.1e) del RGPD, es necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Ley 9/2017 de 8 de noviembre, de Contratos del Sector público	Tramitar electrónicamente los procedimientos de contratación en todas sus fases y posibilitar las relaciones y comunicaciones con licitadores y contratistas. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1 b) y 89.1 del RGPD.
Registro de Colegios Profesionales y Consejos de Colegios de Castilla y León	Consejería de Economía y Hacienda	Secretaría General	Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico :sgeconomiayhacienda@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6,1c) del RGPD , es necesario para el cumplimiento de una obligación legal. Ley 8/1997, de 8 de julio, de Colegios Profesionales de Castilla y León.	Dar publicidad registral a los actos relativos a colegios Profesionales y Consejos de Colegios de acuerdo con la normativa. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD
Gestión de recursos humanos de la Consejería de Economía y Hacienda de la Junta de Castilla y león.	Consejería de Economía y Hacienda	Secretaría General .	Secretaria General C/ José Cantalapidra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico sgeconomiayhacienda@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) del RGPD, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Real Decreto Legislativo 5/2015, de 30 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto Básico del Empleado Público.Ley 7/2005, de 24 de mayo, de la Función Pública de Castilla y León. Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores	Gestión de los recursos humanos de la Consejería de Economía y Hacienda de la Junta de Castilla y León. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD.
Procesos de selección de personal temporal de la Consejería de Economía y Hacienda.	Consejería de Economía y Hacienda	Secretaría General	Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico sgeconomiayhacienda@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) , cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos y 6.1c) del RGPD para el cumplimiento de una obligación legal. Real Decreto Legislativo 5/2015, de 30 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto Básico del Empleado Público.Ley 7/2005, de 24 de mayo, de la Función Pública de Castilla y León. Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores	Procesos de selección de personal interinoy/o laboral eventual para su incorporación a la Consejería de Economía y Hacienda. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD
Controles de acceso y videovigilancia de la Consejería de Economía y Hacienda	Consejería de Economía y Hacienda	Secretaría General	Secretaria General C/ José Cantalapidra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico sgeconomiayhacienda@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) RGPD, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos .Ley 5/2014, de 4 de abril, sde Seguridad Privada. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.	Garantizar la seguridad de personas, bienes e instalaciones, así como el registro y control de acceso de las personas en los edificios de la Consejería Economía y Hacienda.
Registro de altos cargos como miembros de Consejos de Administración en empresas publicas y participadas o de órganos colegiados de Organismos Públicos	Consejería de Economía y Hacienda	Secretaría General.	Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico sgeconomiayhacienda@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) RGPD, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos . Ley 3/2016, de 30 de noviembre, del Estatuto de los Altos Cargos de la Administración de la Comunidad de Castilla y León	Inscripción de los altos cargos que sean miembros de los Consejos de Administración en empresas públicas y participadas de la Comunidad, o de órganos colegiados de organismos públicos. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD
Procedimiento sancionador en Defensa de la Competencia.	Consejería de Economía y Hacienda	Secretaría General.	Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414913. Correo electrónico :defensa.competencia@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) RGPD, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos . Ley 15/2007, de 3 de julio, de Defensa de la Competencia.	Tramitación de los expedientes sancionadores en materia defensa competencia. Denuncias e instrucción de expedientes, recogida de datos de las partes intervinientes y de cualquier tercero con relevancia. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD
Promoción en Defensa de la Competencia.	Consejería de Economía y Hacienda.	Secretaría General .	Secretaria General C/ José Cantalapidra,2.47014 Valladolid. Teléfono 983 414913. Correo electrónico :defensa.competencia@jcyl.es	Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayhacienda@jcyl.es	Artículo 6.1e) RGPD, cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos . Ley 15/2007, de 3 de julio, de Defensa de la Competencia.	Actos de promoción de la competencia dirigidos a prevenir la vulneración de la normativa en esta materia. Los datos personales podrán ser tratados para fines de archivo en interés público, de investigación científica e histórica o fines estadísticos de acuerdo artículos5.1b) y 89.1 del RGPD

Figure 4. Example of published ROPA

2.4.5 GDPR analysis

The OntoROPA project will also review the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).[7]

In particular, it will take into account the regulations specifying obligations related to the Records of Processing Activities (ROPA), art 30, and other related articles such as 31 to 33: "Cooperation with the supervisory authority", "Security of processing", and "Notification of a personal data breach to the supervisory authority".

[7] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

> **Article 30**
> **Records of processing activities**
> 1. Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information:
> (a) the name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer;
> (b) the purposes of the processing;
> (c) a description of the categories of data subjects and of the categories of personal data;
> (d) the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations;
> (e) where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of Article 49(1), the documentation of suitable safeguards;
> (f) where possible, the envisaged time limits for erasure of the different categories of data;
> (g) where possible, a general description of the technical and organisational security measures referred to in Article 32(1).
> […]

This regulation is key for the extraction of knowledge towards ontology design, the validation of competency questions and evaluation of compliance.

2.4.6 Ontology Reuse

As per description in section *2.2.4 Review of GDPR Ontologies* we will review the following ontologies following Ontology Development 101 recommendations and pattern design approaches to establish if it's possible to reuse any or parts of these ontologies and vocabularies.

Table 3. Ontology Reuse

	Formality| Expressivity	**Publicly available**
SPECIAL usage policy language	OWL2	✓
Data Privacy Vocabulary	RDF/OWL	✓
Policy Log Vocabulary	RDF/OWL	✓
DVP-GDPR	RDF/OWL	✓
GDPRov	OWL2	✓
GConsent	OWL2	✓
GDPRtEXT	RDF/OWL(SKOS)	✓
Data Protection Ontology	OWL	✓

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Fiesta-Priv ontology	OWL	✓

2.5 Linked open data

Linked data are essential ingredients of the Semantic Web, where representing information entities via URIs makes them machine processable. The main principles of Linked Data are: entities should be named via unique URIs; these URIs should be HTTP URIs and be resolved using standard web protocols; when these URIs are resolved, they should return useful information about the resource; they should contain links to other URIs so people can discover related resources.

There are many initiatives and research groups focused on work related to the development of vocabularies and terminologies published as open data. For example, in the legal field, Poblet, Casanovas & Rodríguez-Doncel (2019) discuss how to use Linked Data and Open Data in the construction of a model of "linked democracy". Meanwhile, Cimiano et al. (2015) address the advantages of applying Linked Data principles to terminologies and present a model for representing terminologies in RDF.

In Montiel-Ponsoda et al. (2018), the European project Lynx is presented. The project relies on public open data, on the one hand, and on the technologies provided by the Linked Data paradigm, on the other. Lynx, based on a legal knowledge graph, which uses terminologies and thesauri, including the Unesco Thesaurus, shows how to contribute to the construction of advanced services, for example, to annotate documents in the area of Law, to provide definitions of the terms used in them, to classify texts, to identify subjects, and so on.

Moreover, Bosque-Gil et al. (2016), introduce Terminoteca RDF, a prototype that aims to lay down the foundations of a repository of linked multilingual terminologies of official languages in Spain. This project aims to integrate different terminologies into a single unified graph and constitute a single entry point to them. Thus, information coming from different sources and developed in isolation can now be traversed and searched in an easy way by following Semantic Web standards. The core data has been modelled using the Lemon-Ontolex model.

2.6 Innovative ecosystems: Regtech, Lawtech and the transformation of the business model

Innovative ecosystems refer: (i) to the establishment of a framework for the development of vertically scalable knowledge graphs; (ii) the definition of a data model, based on ontologies, for the base definition of the ecosystem graph; and, as we can add in the light of the development of blockchain technologies, (iii) the link of knowledge graphs and data models with decentralised and distributed ledgers to offer security to transactions.

It the last five years, Regtech solutions have been fuelled by the favourable conditions of the legal market. RegTech is an acronym for "Regulatory Technologies". LawTech—regulatory technologies for law—refers to RegTech, FinTech, InsuTech and SupTech. But RegTech is a broader concept, used either in the fields of business, law, management and technology. A simple definition would be "RegTech is about the digital tools that are necessary to master regulatory complexity"(Schäubli,2018). We can distinguish four phases of RegTech development—manual, workflow automation, continuous monitoring, and predictive analytics—mapping services and companies accordingly. We do believe that the technologies of the IoT relate to an upcoming fifth stage, in which sensors will be incorporated to generate a flood of real-time information to be stored, organised and exploited(CbInsights, 2018). The emergence of LawTech web services aims to bring technological solutions and law to business, industry and people, enabling them to better organise and automate both the management of their legal data and legal operations.

LawTech has created an expanding legal market, in which companies offer a variety of legal services mainly based on AI and machine learning solutions—not just the more traditional e-discovery but supervision, monitoring and automatic compliance of regulatory systems, including smart contracts, cryptocurrencies and online dispute resolution. This is a non-complete list of automation fields: (i) expert knowledge and compliance; (ii) legal research (interpretation and resolution of cases), (iii) prediction sentences and cases (legal analytics), (iv) electronic discovery (e-discovery), and (v) intelligent contracts (smart contracts). However, it still is a volatile market. Just before the last pandemic, LawTech venture capital investments increased dramatically at the rate of 2.4 new start-ups per day (Casanovas, 2021). The legal database hosted by CodeX, the Stanford Center for Legal Informatics keeps track of them (Figure 5).[8]

[8] http://techindex.law.stanford.edu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Figure 5. The LegalTech Start-up Landscape, with more than 500 LawTech companies in the USA, Austral-Asia and Europe (Source: Codex, Stanford University).

The automation of legal documents is the most well-trodden path. Legal compliance is the least—as it certainly is a more complex relational field, because the behaviour of all stakeholders must be taken into account (not just meaningful texts to be interpreted).There are systems in legal informatics that have been designed for drafting, storing, organising, consolidating, or retrieving provisions in plain natural language to eventually support legal decision-making (Boella et al., 2013). However, turning norms from natural to formal languages combining NLP techniques and defeasible logic is a difficult task (Wyner et al., 2013). This has not yet been completely solved. The current research is focusing on how to semi-automate the extraction of norms and their elements to populate legal ontologies, combining state-of-the-art general-purpose NLP modules with pre- and post-processing using rules based on domain knowledge to solve the so-called "resource bottleneck problem". Thus, trying to semi-automate the extraction of definitions, norms, and their elements to reduce the need of human intervention(Humphreys et al., 2020). This is a conceptual challenge, lately also called *Rules as Code* in e-government administrations (Waddington, 2020; Governatori et al., 2020).

OntoRopa is benefiting from this expanding market of legal web services. The solution for modelling ROPAs fits into the legal compliance modelling landscape, but we think it is simpler, and easier to be understood, accepted, and adopted not just by LawTech companies, lawfirms and corporations, but by official drafters, rulers, controllers, and supervisors. There is a need to comply with GDPR requirements. Hence, OntoRopa can be expanded through a variety of legal ecosystems, depending on the private or public field of deployment.

Most important, the OntoROPA approach fits nicely into the specific privacy market that will be developed in the European Union in the immediate future.[9] The new strategy mindset represents a shift in the EU's focus, from protecting individual privacy to promoting data sharing as a civic duty. There are initiatives (e.g. the TRUSTS project) to create a pan-European market for personal data through a mechanism called a data trust, a steward that manages people's data on their behalf and has fiduciary duties toward its clients. We do not yet know whether and how this market will be effectively developed, but certainly the solutions provided by OntoROPA are most needed to implement it.

[9] See *A European Strategy for Data* , Brussels, 19.2.2020 COM(2020) 66 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of The Regions. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3 AMBITION: BEYOND THE STATE OF THE ART

3.1 OPPORTUNITIES FOR INNOVATION

3.1.1 State-of-the-art weaknesses and challenges

Our proposal follows this general trend, but instead of interpreting directly the content of article 30 and article 33 of GDPR (mainly about the duties of controllers), we chose an indirect way of approaching the subject:

1) Compliance cannot just be a result to be targeted but a process to be engaged with, embedded into a blockchain solution;
2) Compliance through Design (CtD) means that legal interpretation occurs along the whole process, following several steps crossing hard law, policies, soft law, and ethics;
3) thus, the starting point cannot be a top down nor a bottom up approach, but a middle-out one, stemming from intermediate legal notions—intellectual property, legal time, security, legal validity, etc.— reaching out to all stakeholders involved in the transactions ;
4) in this regard, we are opening a legal procedural way of producing veracity, certainty and especially trust, as consumers, producers and markets will have a mechanism to turn out their Records of Processing Activities (ROPAs) into a legal, acceptable and actionable document;
5) what this latter formulation entails is a certification process that can be accepted by agencies and courts as legal evidence, turning the needle in all directions of the legal compass (Fig. 1);
6) in this sense, we do not need to wait for a specific case-based interpretation of what 'joint controller' means (there are no available cases yet): stemming from the notions and clusters contained into the documents produced by Data Protection Agencies should be enough to get a good description of official implementation patterns;
7) therefore, as said, the knowledge acquisition process (KAP) should start from the documents and the actual behaviour already in place, and not from any abstract interpretation of how the process should be;
8) the final result of the project lifecycle is a certified ROPA that can be offered as a legal web service on the iExec platform.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3.1.2 Privacy compliance: What ROPA inform about and why it is important in the European data economy

The value of legal compliance in business processes has been shown in section 2.6. Legal compliance includes privacy compliance, which in Europe is regulated by the General Data Protection Regulation (GDPR) Regulation (EU) 2016/679) (EU 201657).

GDPR regulation enforces the obligation of any agent (public or private entity)who treats personal data to keep a Register **(ROPA)**. ROPA´s content is regulated in Article 30 of the GDPR

ROPAs are indeed very important pieces of information.There is no other tool in the GDPR to cover the recording of what is being done with personal data.

Whenever a data breach occurs the Data Protection supervisors will always ask for this content to be made available for inspection.

ROPAs are also key elements of legal smart governance, as they are the tool we citizens are provided with for democratically access to the information on how our personal data is (or not) treated and protected.

And they can play an even more important role: they can sustain the **evidence** that authorities (data protection supervisors, judges, etc..) accept as proof of being proactive in privacy compliance **(compliance through design)**.

ROPA shares information (i) about the person or organization responsible of the treatment if personal data travels outside the European Union,(ii) and about the security measures taken to protect them. However, it is really easy to find processes that treat personal data and no ROPA is present.

Even if ROPAs are available, *how can we be certain that a ROPA contains the information it should*? Who creates it?

Currently, DPOs (Data Protection Officers) are in charge of this task, which they accomplish by inspecting ROPAs manually.

3.1.3 How can semantic interoperability help privacy compliance to go beyond information?

The Ontoropa team is in addition very interested in the issue of the ***lack of interoperability*** in current ROPAs. Most of these ROPAs are created and published as pdf or excel files.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

They are not interoperable. Neither from a syntaxis perspective or a semantic perspective. It is impossible to apply IA methods or other methods that are able to infer new knowledge from data if *semantics is not available* for automatic processing.

3.1.4 Privacy compliance and distributed computing: conflicts and challenges

The use of some technologies, such as blockchain, entails the existence of several possibilities to vulnerate the GDPR (CNIL, 2018). The distributed nature of blockchain technology and the anonymity that characterizes it is not well aligned with the transparency that GDPR requires about the "who" (who is responsible, who does the treatment, who is responsible when a privacy breach occurs) and the "where" (personal data should stay in areas where citizenship rights are protected in accordance with EU privacy regulations).

- But how many data processors that use blockchain are treating these issues in the right way?
- What are they doing to comply with GDPR?
- What do they do to prevent data transfers that should not be made, or to erase personal data when there is no longer any reason to justify keeping them?
- Are they creating the appropriate ROPAs? If they are not, they do not comply, They are subject to a fine of up to 4% of their annual income (GDPR).

3.1.5 Proactiveness: the Achille's heel of privacy compliance

GDPR involves the evolution from a reactive model to a proactive model. Reacting and complying after the issue is no longer supported. This behavior is not GDPR compliant and subject to sanctions.
Therefore, it is necessary to find new ways to demonstrate that proactivity in legal compliance is accurate. How can we demonstrate that a ROPA has been created since the beginning of the personal data treatment? How can it be shown that a ROPA is not false but has been created by the stakeholder who must do so? How can it be shown that a ROPA is credible?

3.1.6 Knowledge sharing for research progress

How can the information that ROPAs contain be used to obtain *knowledge that can help to improve the state of the art of privacy compliance and security technologies?*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

For example, if the information ROPA contain about security measures would be accessible and ready to be treated in inference processes, security researchers could know what security measures are more popular, or more effective for protecting personal data in public networks, what platforms provide the better tools for this aim, etc. To obtain such type of knowledge, ROPAs should no longer be isolated pieces of information. They should be reliable sources of information, linked, available for intelligent knowledge extraction.

3.2 GOALS

1) To represent ROPAs with knowledge graphs that can be assessed with automatic and intelligent processes. This will be supported by a professional standard ontology for ROPAs, the OntoROPA ontology, and Semantic Web technologies.

2) To provide support for sharing and linking information abou ROPAs in a community-based ecosystem, ready to be exploited by automatic processes.

3) To support proof of proactiveness with automatic processes.

3.3 WHY IS THIS INNOVATIVE?

- Technology will be used to proof proactiveness in privacy compliance. To our knowledge, there is no similar solution. The proof will be automatic, irrefutable.

- The correctness of ROPAs content, which is a part of legal compliance (a ROPA must contain what GDPR requires), will be validated with automatic processes. Again, there is no similar solution.

- Treatments that use blockchain technology will get help to recognize situations that require to create a ROPA.

- To our knowledge, there is no ontology such as the one we intend to obtain. There are ontologies that model the GDPR (1) and legal ontologies (2), but not specific for ROPAs. Moreover, we are aware of the existence of two software tools that help creating ROPAs (3,4). However, none of them seem to be built on top of an ontology, as they do not offer any reference to such type of knowledge tool. However,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

our intention is to create an ontology that will be offered as open data, an OWL file, reliable, reusable, and extensible.

While some of challenges in section 3.3 have similarities to the challenges encountered for different communities, such as workarts and NFTs, it is worth remarking that these challenges are novel in the area of legal compliance.

Legal compliance introduces the requirement of *evidence* and *responsibility*, therefore the challenge to provide mechanishms that can *be accepted by legal authorities* (judges, legal compliance supervisors). Providing such type of *reinforced* mechanisms, for such a demanding environment, is innovative.

3.4 SCENARIOS

3.4.1 Scenario 1: Proof (certification) of privacy compliance using ONTOCHAIN ecosystem

In this scenario the original ROPA is created outside the ONTOCHAIN ecosystem. It applies to existing ROPAs, and ROPAs that can be created in the future with similar methods, or with a software tool for ROPA editing. Current ROPAs are indeed created with artisanal methods, resulting in .pdf or .xls files. Tranformation of ROPAs already available as .xls files is the Use Case 1. The production of ROPAs from start with a ROPA management tool is Use Case 2. Figure 1 shows the process flow for use case 1. Figure 2 shows the data flow.

Our solution produces a RDF ROPA, with semantic annotations. This ontological ROPA has two properties:

- The ROPA is verified. Its content is validated against Article 30 of GDPR. This is achieved by using the professional ontology, OntoROPA.
- The ROPA is certified. Once verified, it is registered and freezed as proof of proactiveness.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

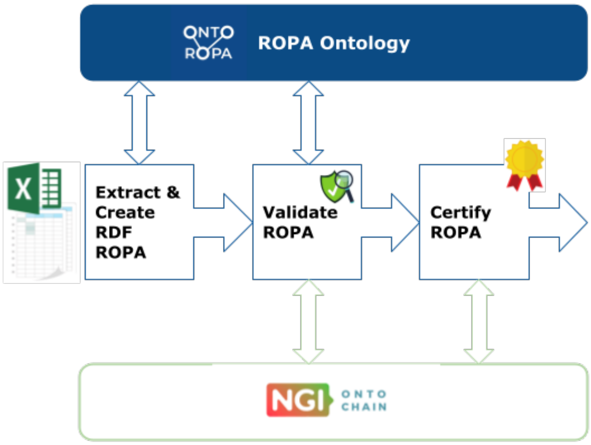


Figure 6. Process flow for use case 1, source is a .xls ROPA.

3.4.2 Scenario 2: Privacy compliance through design in ONTOCHAIN ecosystem

This scenario is related to the blockchain difficulties with privacy compliance. It deals with treatments that use blockchain technology to treat/store personal data. A use case is a service offered in the ONTOCHAIN ecosystem that processes personal data in the blockchain. No matter how it protects these data or whether it uses the best technical facilities towards this aim: a ROPA must be created, as it is mandatory. So, we will consider a general process which can easily represent any of the projects that deal with personal data in this Call (in section 4 some of them listed). It is worth noting that the main difference of this use case with the use cases of scenario 1 is the role on the ONTOCHAIN ecosystem. In scenario 1, the ROPA is created out of the ONTOCHAIN ecosystem, which is used to provide trustworthiness. In scenario 2, the ONTOCHAIN ecosystem is enriched with the capability to improve privacy compliance by default.

In this scenario the use case starts with an analysis assessing whether the treatment started in ONTOCHAIN needs or does not need, a ROPA. The flow of this decision is shown in figure 3. If the treatment requires a ROPA to demonstrate privacy compliance, the output is a recommendation about this need, and a suggestion for ROPAs content.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

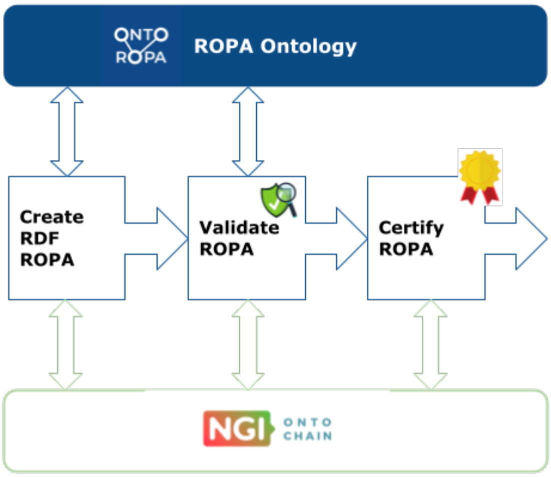


Figure 7. Scenario 2, privacy compliance through design in the ONTOCHAIN ecosystem.

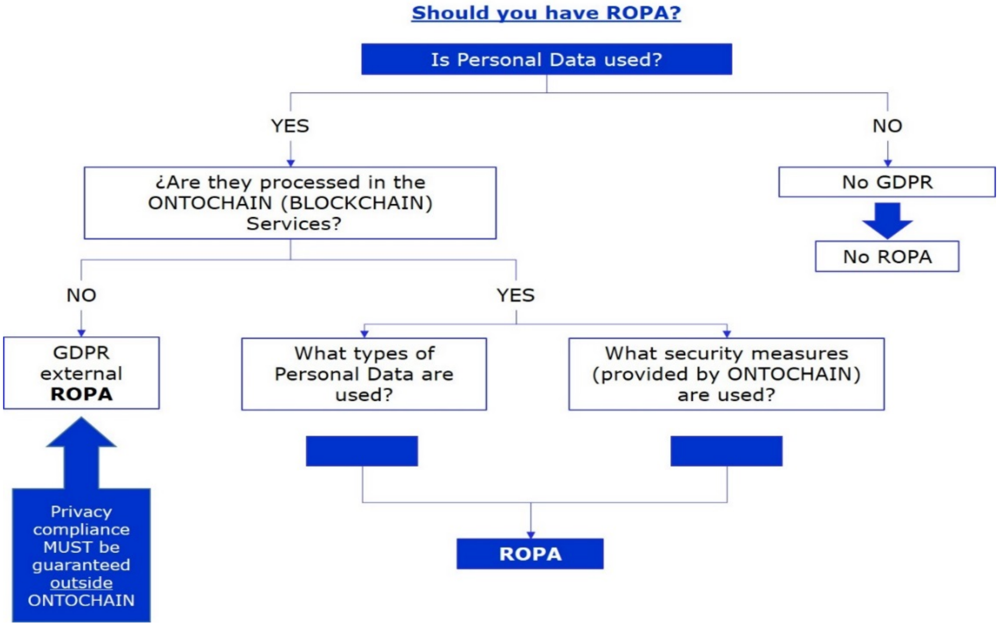

Should you have ROPA?
Is Personal Data used?
YES
NO
¿Are they processed in the ONTOCHAIN (BLOCKCHAIN) Services?
No GDPR
No ROPA
NO
YES
GDPR external ROPA
What types of Personal Data are used?
What security measures (provided by ONTOCHAIN) are used?
Privacy compliance MUST be guaranteed outside ONTOCHAIN
ROPA

Figure 8. Profiling (classification) of ONTOCHAIN treatments in Scenario 2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

4 SYNERGIES WITH ONTOCHAIN AMBITIONS AND ADDED VALUE FOR ONTOCHAIN

4.1. ONTOROPA IN ONTOCHAIN ECOSYSTEM

The OntoROPA proposal fits within the "Metadata and document assessment" ONTOCHAIN use case. OntoROPA provides metadata about ROPAs that can be assessed with automatic and intelligent processes.

4.1.1 OntoROPA and ONTOCHAIN challenges

OntoROPA fits within the ONTOCHAIN challenges collected in topic 2, "SEMANTIC INTEROPERABILITY". The most relevant alignments appear in Table 4.

Table 4. Alignment of ONTOCHAIN challenges with OntoROPA challenges

ONTOCHAIN challenge	OntoROPA challenge
Develop a trustworthy, privacy-preserving, secure, transparent, democratic and traceable **approach to manage access and operations over ontologies, metadata, data, knowledge and information in the ONTOCHAIN ecosystem.** Particularly, provide practical solutions that rely on already successful Semantic Web approaches such as Linked Data, OWL Lite, OWL DL and other approaches and formats.	To develop a trustworthy and transparent approach to manage access over ontologies, metadata, knowledge and information, providing practical solutions that rely on already successful Semantic Web approaches such as Linked Data and OWL
The research and innovation should also cover solutions for validation that ontology and on- chain and off-chain knowledge is logically consistent, or alerts in cases of inconsistencies.	To offer ontology-based solutions for validation that ROPAs are logically consistent.
Development of specific ontologies to be used for the project use cases for each application domain, … certification, and the collection and maintenance of various ontologies useful for the project applications and use cases	Development of a professional ontology, the OntoROPA ontology, that collects the expertise of professionals in charge of ROPA creation and maintenance. Cerification that ROPA life cycle is GDPR compliant.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

4.1.2 OntoROPA and ONTOCHAIN requirements

OntoROPA is aligned with ONTOCHAIN requirements of open data, semantic interoperability and distributed knowledge sharing. Alignment of OntoROPA characteristics with ONTOCHAIN requirements is presented in Table 5.

Table 5. OntoROPA and ONTOCHAIN requirements

ONTOCHAIN requirement	OntoROPA
Design specific domain ontologies	OntoROPA ontology is an ontology that collects expertise from professionals dealing with ROPAs
Trustworthy information and knowledge management operations for content, services, clusters, hierarchies or similar	ROPAs will be certificated in order to provide the trustworthiness that will be proof of legal compliance. Reasoning on top of the ontology will allow for ROPAs comparison and extraction of new knowledge.
Validate the correctness on ontology data instance via blockchain	ROPAs content will be validated in secure environments, able to guarantee and proof that the process is executed in a secure trusthworthy manner

4.1.3 OntoROPA synergies with ONTOCHAIN architectural components

Regarding the initial ONTOCHAIN architecture with components at different levels—Distributed Leger, Core Protocols, Application Protocols, Use-Cases, and Solution Domain. We point out the most interesting ones for OntoROPA:

- Identity Management/Identification: Legal compliance requires being able to link responsibilities and authorhisp to legal entities, real world entities. OntoROPA will use identity management to do it.
- Data Semantics/Semantic Linking: the OntoROPA project aims to represent ROPA as RDF graphs, linked with the ONtoROPA ontology, but also to other ROPAs. RDF, linked data and related Semantic Web standards provide the tools to represent, share and manage semantics in technical environments.
- Smart Contracts. Transactions associated to ROPA verification, certification and publication can be expressed as smart contracts.
- Data provenance. ROPAs are changing documents, they are maintained and they should mirrow the updates that real personal data activities suffer. It is necessary to trace ROPAs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

- Market mechanisms. ROPAs can themselves be input data for research scientists who need smart data for their research. Moreover, open access to ROPAs is not a universal right granted by GDPR. Restrictions can be stablished, and this access can be granted through policies that can be modeled as market policies.
- Trustworthy information exchange. Talking about trustworthy and smart ROPA means talking about exchanging trustworthy information. It is not enough to have smart data as input. It is necessary that exchanges are also trustworthy.
- Trustworthy web and Community-based Knowledge Sharing. This is one of the open possibilities shown in figure 1 for the highest layer, Solution Domain. It is the community who manages and accesses them to obtain information which sustains the interest of trustworthy ROPAs. The interest of some members of the community to makes sure that the information they access, the ROPAs, are trustworthy is the motivation for ROPA certification and validation. In reciprocity, there is the interest of the other members of this community, ROPA maintainers, in proving that the information they provide is trustworthy, and that they comply with privacy rules.

OntoROPA can benefit from the projects which are able to provide innovative methods as follows:

- **Certification of the genesis or origin of a ROPA.**
  Who has created the ROPA, publishes it, and therefore responds to the authorities? For example, when a security incident occurs with personal data, such as a data breach or data leak, the data protection authority may require ROPAs to verify the processing and security measures taken in each of them.

  The entity in charge of a ROPA, which will in turn be in charge of the data processing, is responsible for creating it and making it public. This is the case of Spanish Public Administrations, which hold enforcing powers and can take punitive actions when an incident occurs.

  Since we are talking about legal responsibility, it is essential to associate authorship with a legally recognized identity, i.e. a natural or legal person. Users of technology platforms are not valid entities.

- **Certification of the quality of ROPAs demonstrate that its content corresponds to what the GDPR requires.**
  Validating a ROPA means verifying that it actually contains what the GDPR requires.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

If the verification, which will necessarily use a professional ontology, produces a satisfactory result (valid ROPA), it will be also necessary to certify that it has been successfully validated by the OntoROPA algorithm.

- **Proof of proactivity, which in the case of these records means being able to prove that they were created when required by the GDPR, that is, when the processing of personal data described in them was initiated.**

  These requirements have some similarity with some of the FTTs, although in the case of ROPAs a differentiating and fundamental element is introduced, which is legal compliance in its broadest sense. It is seeking assurance, a proof that it is accepted as socially valid and supported by the authorities. In the end, it guarantees rights, it can be used for conflict resolution and, as a result, it adds value.

  Identity management, proactivity testing, and semantic validation testing are innovative elements that differentiate NFT ROPA.

What OntoROPA offer to other projects?

1) OntoROPA is a true use case, based on the experience of professionals working with ROPAs, with an approach that captures real needs of our society. This makes it an excellent use case for other projects that propose innovative technological solutions within ONTOCHAIN.
2) In addition, OntoROPA can help projects that handle personal data detecting the need for ROPA to comply with privacy requirements and create such ROPA if necessary.

4.2. SYNERGIES WITH SPECIFIC TEAMS

All projects submitted to ONTOCHAIN have been analyzed and a series of interviews have been arranged over the past week, which have allowed us to classify synergies into four fundamental blocks:

1.Identity

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

DR HIBI

- The team has experience and solutions for identity management using blockchain.
- Collaborating with Datarella to manage the identity of organizations would facilitate a solution for the identity management of legal entities that OntoROPA needs.

OntoSSiVault

- The team offers solutions for digital identity management. They have solutions for identity management applied to corporates.
- Their solutions for identity management applied to corporates are related to OntoROPA´s goal of advancing the state of the art of legal compliance by providing tools capable of certifying the identity of the legal entity responsible for a ROPA.In many situations, these legal entities are corporates, universities, schools, public administrations…

2.Processing: Storage and certification of knowledge networks

GraphChain

- This team provides a security and trust system adapted to RDF graph.
- Since OntoROPA proposes the use of knowledge graphs represented with RDF as the data model to work on, OntoROPA could use GraphChain solutions to store and certify the validity of its knowledge graph.
- GraphChain could provide an interface, using serialized RDF ROPA files can be injected into GraphChain.

3.Secure Execution

KX-KnowledgeX

- The team offers solutions for trusted executions.
- Their solution could execute data processing of user-related data in a TEE and generate certifications.
- KX-KnowledgeX enable end user to become "raw data owners" and audit how their data was processed.
- Architecturally, OntoROPA can use KnowledgeX as the technical vehicle to enable traceability and auditability of data.

4.Ecosystem for a (social) community:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

SEIP

- This team offers creating virtual blockchain on top of public blockchain. This virtual community is enabled by access policies.
    - OntoROPA can benefit from the collaboration with SEIP project for community management. OntoROPA proposes for Phase 2 (see the Proposal) a community-based collaboration to evolve the ontology prepared for the Proof of Concept to a standard. During discussions and the preparation of the standard, interactions of this community would benefit from the facilities seIP provides in establishing policies.

We underline that the discussions held are the first step for possible future collaborations, especially for those teams interested in our use case.

Likewise, the round of interviews has allowed us to detect, in addition to the clear collaborative aspects mentioned in the previous point, the interest in the help that OntoROPA can provide to projects that handle personal data. Specifically, TENACIOUS Project, a team that offers semantic services in the cloud, expressed its interest in exploring the possible use of a module/tool that would allow verifying whether a service provided by **TENACIUS** would require a semantic ROPA to ensure legal compliance with respect to the GDPR. The next meeting will serve to deepen this proposal.

Finally, it has not been possible to complete all the planned interviews with the teams, but we hope to be able to carry them out in the next few days.

Benefit of synergies between OntoROPA and other projects for the ONTOCHAIN ecosystem:

1) Being one of the first ecosystems that enable GDPR compliance for user-related data can be a game changer. Until this very day the legal principles are established, but their implementation and enforcement are in many cases either not existing or suboptimal.

2) The synergy between the referred teams and OntoRopa will ensure a trusted execution of knowledge generation, and can have a larger impact on DGPR compliance as it will secure the process of legal validation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

5 CONCLUSIONS

OntoROPA provides solutions for smart legal compliance. We have demonstrated the innovation opportunities offered by the current state of the art in legal compliance. OntoROPA works on them.

Since the beginning of this call, OntoROPA has defined two different scenarios and advanced synergies with other ONTOCHAIN projects. In the first scenario, other projects can provide OntoROPA with solutions to achieve some certification, validity, and trust requirements. In the second one, OntoROPA helps other ONTOCHAIN projects to achieve legal compliance.

Innovation opportunities in the legal compliance market, including blockchain-based solutions, are excellent.Collaboration with other projects will be of great benefit for OntoROPA, and to the ONTOCHAIN ecosystem.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

REFERENCES

Akhavan, P., Shahabipour, A., & Hosnavi, R. (2018). A model for assessment of uncertainty in tacit knowledge acquisition. *Journal of Knowledge Management*, *22*(2), 413–431. https://doi.org/10.1108/JKM-06-2017-0242

Akhigbe, O., Amyot, D., & Richards, G.(2015). Information technology artifacts in the regulatory compliance of business processes: a meta-analysis. In M. Benyoucef et al. (eds.), *MCETECH 2015* (pp. 89–104). LNBIP 209.

Akhigbe, O., Amyot, D & Richards, G (2019). A systematic literature mapping of goal and non-goal modelling methods for legal and regulatory compliance. *Requirements Engineering, 24*(4), 459-481.

Boella, G., Tosatto,S.C., Ghanavati, S., Hulstijn, J.,Humphreys, L.,Muthuri, R., Rifaut, A., & van der Torre, L.(2013). Integrating legal-urn and eunomos: Towards a comprehensive compliance management solution. In *International Workshop on AI Approaches to the Complexity of Legal Systems* (pp. 130-144). Springer.

Bosque-Gil, J., Montiel-Ponsoda,E., Gracia, J., & Aguado-de-Cea,G.(2016. Terminoteca RDF: a Gathering Point for Multilingual Terminologies in Spain, in *Term Bases and Linguistic Linked Open Data. Proceedings of TKE 2016 the 12th International conference on Terminology and Knowledge Engineering* (pp. 136–146). Copenhagen Business School.

Campanile, L., Iacono, M., Marulli, F. and Mastroianni, M.(2021). Designing a GDPR compliant blockchain-based IoV distributed information tracking system. *Information Processing & Management*, 58(3), 102511.

Casanovas, P. (2021). Inteligencia Artificial y Derecho: La doble implosión de las profesiones y servicios jurídicos en la era digital" [Artificial Intelligence and Law: The Double Implosion of Legal Professions and Services in the Digital Age]. In Velarde, Olivia & Martín, Manuel (Eds.) *Mirando hacia el futuro. Cambios sociohistóricos vinculados a la virtualización*. Centro de Investigaciones Sociológicas (CIS, Ministerio de la Presidencia. In Press.

Casanovas, P., González-Conejero, J., & de Koker, L.(2017). Legal compliance by design (LCbD) and through design (LCtD): preliminary survey. http://ceur-ws.org/Vol-2049/05paper.pdf.

Casanovas, P., Hashmi, M., Lam, B.,& de Koker, G.(2021). *Legal Compliance by Design (LCbD) and through Design (LCtD): A Literature Survey* (forthcoming).

Casellas, N. (2011). Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge. Springer.

CbInsights. (2018). Regtech 102: The Evolution Of Regtech And The Future Of Regulatory Compliance, January 9, https://www.cbinsights.com/research/regtech-four-phases-expert-intelligence/

Cimiano, P. et al. (2015). Linked Terminologies: Applying Linked Data Principles to Terminological Resources. *Proc. Fourth Bienn. Conf. Electron. Lexicogr* (pp. 504–517).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

CNIL (2018). *Blockchain. Solutions for a responsible use of the blockchain in the context of personal data.* https://www.cnil.fr/sites/default/files/atoms/files/blockchain_en.pdf

Creswell, J. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Data Privacy Vocabulary -DPV (2021). *Data Privacy Vocabulary (DPV)* (Version 0.2). Draft Community Group Report, 13 January 2021. https://dpvcg.github.io/dpv-gdpr/

Dehghani, M., & Akhavan, P. (2017). An experimental investigation of knowledge acquisition techniques. Journal of Management Development, 36(4), 493-514. https://doi.org/10.1108/JMD-07-2016-0132

Esteves, B., & Rodríguez-Doncel, V. (2021 forthcoming). Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR, submitted to *Semantic Web Journal* (2703-3917). http://www.semantic-web-journal.net/system/files/swj2703.pdf.

European Union Blockchain Observatory & Forum. (2020). *Report on EU Blockchain Ecosystems Developments, 20 November 2020.* https://www.eublockchainforum.eu/sites/default/files/reports/EU%20Blockchain%20Ecosystem%20Report_final_0.pdf

Faber, B., Michelet, G.C., Weidmann, N., Mukkamala, R.R. & Vatrapu, R.(2019. BPDIMS: A blockchain-based personal data and identity management system. In *Proceedings of the 52nd Hawaii International Conference on System Sciences.*

Fink, M.(2019)*Blockchain and the General Data Protection Regulation. Can distributed ledgers be squared with European data protection law?* EPRS, European Parliamentary Research Service.

Scientific Foresight Unit (STOA), PE 634.445 – July (2019). https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf

Floridi, L. & Cowls, J.(2019). A unified framework of five principles for AI in society. *Harvard Data Science Review.* 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Governatori, G., Idelberger, F., Milosevic, Z., Riveret, R., Sartor, G. & Xu, X. (2018). On legal contracts, imperative and declarative smart contracts, and blockchain systems. *Artificial Intelligence and Law*, 26(4),377-409.

Governatori, G., Barnes, J., Zeleznikow, J. Hashmi, M., de Koker, L.,Poblet, M., & Casanovas, P. (2020). `Rules as Code' will let computers apply laws and regulations.But over-rigid interpretations would undermine our freedoms. *The Conversation*, 26 November 2020, https://theconversation.com/rules-as-code-will-let-computers-apply-laws-and-regulations-but-over-rigid-interpretations-would-undermine-our-freedoms-149992

Hashmi, M., Casanovas, P., & de Koker, L.(2018). TERECOM2018@ JURIX, Technologies for Regulatory Compliance "Legal Compliance Through Design: Preliminary Results of a Literature Survey". *TERECOM2018@ JURIX, Technologies for Regulatory Compliance.* http://ceur-ws.org/Vol-2309/06.pdf

Hashmi, M., Governatori, G., Lam, H.P. & Wynn, M.T.(2018). Are we done with business process compliance: state of the art and challenges ahead. *Knowledge and Information Systems*,57(1),79-133.

Humphreys, L., Boella, G., van der Torre, L., Robaldo, L., Di Caro,L., Ghanavati, S., & Muthuri, R. (2020). Populating legal ontologies using semantic role labeling. *Artificial Intelligence and Law*, 1-41.

Kirrane, S., Villata, S., & d'Aquin, M. (2018). Privacy, security and policies: A review of problems and solutions with semantic web technologies. Semantic Web, 9(2), 153-161.

Kost, M., & Freytag, J. C. (2012). Privacy analysis using ontologies. In E. Bertino, & R. S. Sandhu (Eds.), *2nd. ACM Conference on Data and Application Security and Privacy, CODASPY 2012* (pp. 205-216). ACM.

Kost, M., Freytag, J. C., Kargl, F., & Kung, A. (2011). Privacy verification using ontologies. In *Sixth International Conference on Availability, Reliability and Security, ARES 2011* (pp. 627-632). IEEE Computer Society

Montiel-Ponsoda,E.,Rodríguez-Doncel, V., Martín-Chozas, P., & Kernerman,I. (2018). Lynx and the Legal Knowledge Graph: Integrating lexical and terminological resources with legal data. *Kernerman Dict. News*, 261565-47 (26), 2-5.

Neuman, W. (2011). *Social research methods: qualitative and quantitative approaches* (7th ed.). Allyn & Bacon.

Palmirani, M., & Governatori, G. (2018). Modelling legal knowledge for GDPR compliance checking. *Frontiers in Artificial Intelligence and Applications*, 313, 101-110. https://doi.org/10.3233/978-1-61499-935-5-101

Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., & Robaldo, L. (2018). Legal ontology for modelling GDPR concepts and norms. *Frontiers in Artificial Intelligence and Applications*, 313, 91-100. https://doi.org/10.3233/978-1-61499-935-5-91

Pandit, H.J.(2020). *Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance.* [PhD Thesis]. Trinity College Dublin. https://harshp.com/research/publications/035-representing-activities-processing-personal-data-consent-semweb-gdpr-compliance

Poblet, M., Casanovas, P., & Rodríguez-Doncel,V.(2019). *Linked Democracy*. Springer.

Poblet, M., Casellas, N., Torralba, S., & Casanovas, P. (2009). Modeling expert knowledge in the mediation domain: a middle-out approach to design ODR ontologies. In *Legal Ontologies and Artificial Intelligence Techniques. Workshop on Legal Ontologies and Artificial Intelligence Techniques* (No. 3è).

Rodrigues, C. M. de O., Freitas, F. L. G. de, Barreiros, E. F. S., Azevedo, R. R. de, & de Almeida Filho, A. T. (2019). Legal ontologies over time: A systematic mapping study. *Expert Systems with Applications*, 130, 12-30. https://doi.org/10.1016/j.eswa.2019.04.009

Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (Eds.).(2011). *Approaches to Legal Ontologies*. Springer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Schäubli, T.(2018). RegTech Definition: How to Define Regulatory Technology, July 7 2018. https://blog.apiax.com/regtech-definition-how-to-define-regulatory-technology-c8d2e5569852

Schwerin, S. (2018). Blockchain and Privacy Protection in the Case of the European General Data Protection Regulation (GDPR): A Delphi Study.*The Journal of The British Blockchain Association, 1*(1), 23-31.

Waddington, M. (2020). Research Note. Rules as Code. *Law in Context*, 37(1), 1-8. https://doi.org/10.26826/law-in-context.v37i1.134

Wiśniewski, D., Potoniec, J., Ławrynowicz, A., & Keet, C. M. (2019). Analysis of Ontology Competency Questions and their formalizations in SPARQL-OWL. Journal of Web Semantics, 59, 100534. https://doi.org/10.1016/j.websem.2019.100534

Wyner, A.Z., & Governatori, G.(2013). A Study on Translating Regulatory Rules from Natural Language to Defeasible Logics. In *RuleML* (2).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

APPENDIX A

Table 1. Ontology Methodologies

Methodology	Year	Preparatory Step	Development Step	Evaluation Step
Grüninger & Fox King, Grüninger & Uschold	1995	1. Specify scenario 2. define requirements	3. define terminology 4. Specify definitions and constraints	5. Evaluate with CQ
King, Grüninger & Uschold	1995/ 1996	Purpose	Ontology building: - Capture - Coding - Integrating	Evaluation
Methontology[10]	1997	Specification	Conceptualization -glossary -taxonomies -relations -concept dictionaries -axioms & rules Formalization	Evaluation
Ontology Development 101[11]	2001	1. Domain & scope 2. Reuse	3. Term list 4. Classes 5. Class properties 6. Slot facets 7. Instantiate	
OTK Methodology[12]	2002 (1993)	Feasibility study Kickoff (ORSD)	Kickoff -source analysis -create draft ont. Refinement -**KA experts** -draft modification -formalization	Evaluation
DILIGENT			1 build	

[10] Fernandez-López, M., Gomez-Perez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In Farquhar, A. & Gruninger, M. (Eds.), *Ontological Engineering: Papers from the 1997 Spring Symposium AAAI97, volume Technical Report SS-97-06* (pp. 33-40). American Association for Artificial Intelligence.

[11] Noy, N. F. & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology. Technical Report SMI-2001-0880.* Stanford University School of Medicine.

[12] Sure, Y., Staab, S., & Studer, R. (2002). Methodology for development and employment of ontology based knowledge management applications. *ACM SIGMOD Record*, 31 (Special Issue)(4), 18-23.
Sure, Y. & Studer, R. (2002). On-to-knowledge methodology - final version. *Project Deliverable D. 18, EU IST-1999-10132 On-To-Knowledge.* Institute AIFB, University of Karlsruhe.
Sure, Y. & Studer, R. (2003). A methodology for ontology-based knowledge management. In Davies, J., Fensel, D., & van Harmelen, F. (Eds.),*Towards the Semantic Web. Ontology-driven Knowledge Management* (pp. 33-46). John Wiley & Sons.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Methodology	Year	Preparatory Step	Development Step	Evaluation Step
			2 Local adaptation 3 Analysis 4 Revision 5 Local update	
NeOn[13]	2007	1. Purpose, scope and level of formality 2. Users 3. Uses 4. Requirements CQ	5. Group requirements 6. Validate Requirements 7. Prioritize Requirements 8. Extract terms	Evaluate against ORSD
UPON[14]	2009	Requirements W. -domain -purpose -scope -CQ	Analysis W. -reuse -term extraction -concept definition -relations -UML Implementation	Test -consistency -relevance -completeness
Ontology Design Patterns[15]	2009	- CQ	Method: - Reengineering from other data models - Specialization/Composition of other CPs - Extraction from reference ontologies - Creation by combining extraction, specialization, generalization, and expansion	
MeLoN Methodology for building Legal Ontology)[16]	2019	1. Describe the goal of the ontology (storytelling) 2. Evaluation indicators and parameters/indicators to	4. List all the relevant terminology and produce a glossary 5. Use tables to model the knowledge-base of	10. Evaluate the ontology using the OntoClean method and 2) 11. Publish the document with the LODE tool and github

[13] Espinoza, M., & Sabou, M.(2008). *Neon methodology for building contextualized ontology networks. Deliverable D5.4.1, NeOn Project*. http://www.neon-project.org.

[14] de Nicola, A., Missikoff, M., & Navigli, R.(2009). A software engineering approach to ontology building.*Information Systems*, 34, 258-275. http://wwwusers.di.uniroma1.it/~navigli/pubs/De_Nicola_Missikoff_Navigli_2009.pdf

[15] Gangemi, A. & Presutti, V. (2009). *Ontology Design Patterns*. In *Handbook on Ontologies* (pp. 221-243). 10.1007/978-3-540-92673-3_10.

[16] https://decodeproject.eu/file/718/download

Methodology	Year	Preparatory Step	Development Step	Evaluation Step
		evaluate the ontology 3. State of the art survey and other existing domain vocabularies	the legal domain (excel) 6. Transform the tables in UML model using the Graffo tool 7. Transform the UML into OWL/XML serialization 8. Test the output under the technical and legal point of view (SPARQL queries on individuals) 9. Refine and optimize OWL by ontologist experts	12. Collect feedbacks from the community and validation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

APPENDIX B

SPECIAL Usage Policy Ontology and Related Vocabularies

The European H2020 project SPECIAL (Scalable Policy-aware Linked Data Architecture for Privacy, Transparency and Compliance) aimed at providing technical solutions for data protection requirements associated to big data. During this project several ontologies and vocabularies were developed.[17]

The **Usage Policy ontology** (SPECIAL usage policy language), was developed to represent data processing activities.

> *This document specifies the SPECIAL usage policy language, which can be used to express both the data subjects' consent and the data usage policies of data controllers in formal terms, understandable by a computer, so as to automatically verify that the usage of personal data complies with data subjects' consent.*[18]

Several related vocabularies were also developed: Data Privacy Vocabulary (DPV) and its extension DVP-GDPR and the Policy Log Vocabulary.[19]

> *The Data Privacy Vocabulary (DPV) provides terms (classes and properties) to describe and represent information related to processing of personal data. This extension [DVP-GDPR] extends the DPV and provides concepts specific to the obligations and requirements of the General Data Protection Regulation (GDPR). More specifically, it provides a taxonomy of legal bases and rights as defined within the GDPR.*[20]

GDPRtEXT, GDPRov and GConsent

The PhD research of Harshvardhan J. Pandit contributed to this domain with two semantic web ontologies:

[17] Bonatti, P. A, Bos, B., Decker, S., Fernandez,J. D., Peristeras, V., Polleres, A., & Wenning, R. (2018). Data Privacy Vocabularies and Controls: Semantic Web for Transparency and Privacy. In *Proceedings of the Workshop on Semantic Web for Social Good Co-Located with 17th International Semantic Web Conference (ISWC 2018)*, 4. http://ceur-ws.org/Vol-2182/paper_3.pdf.
Bonatti, P. A., Kirrane, S., Petrova, I. M., Sauro, L. & Schlehahn, E. (2018). *D2.5 Policy Language V2. Scalable Policy-awarE Linked Data arChitecture for prIvacy, transparency and compLiance (SPECIAL)*.https://www.specialprivacy.eu/images/documents/SPECIAL_D25_M21_V10.pdf.
Bonatti, P. A., Ioffredo, L., Petrova, I. M., Sauro, L., & Siahaan, I. R. (2020). Real-time reasoning in OWL2 for GDPR compliance. *Artificial Intelligence*, 289. https://doi.org/10.1016/j.artint.2020.103389
[18] https://ai.wu.ac.at/policies/policylanguage/
[19] https://www.specialprivacy.eu/platform/ontologies-and-vocabularies
[20] https://dpvcg.github.io/dpv-gdpr/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

- **GDPRov** represents information about activities associated with processing of personal data and consent. The GDPRov is an extension of the P-Plan ontology that uses PROVO's prov:Plan to model expected workflows.
- **GConsent** represents information associated with determining compliance of consent (concept management).
- **GDPRtEXT** is a resource that provides a linked data version of the GDPR text and a glossary of its concepts. Both ontologies define concepts and relationships using GDPRtEXT to indicate source within GDPR.

> *Together with GDPRtEXT, GDPRov and GConsent enable representation of activities required to evaluate and validate compliance with the GDPR. Apart from advancing state of the art, the ontologies also provide a vocabulary of terms and concepts relevant for GDPR compliance, and demonstrate the use of legal documents as a source for ontologies using linked data principles.*[21]

GDPRov and GConsent are published under an open license (CC-by-4.0).[22]

Data Protection Ontology and The DAta Protection REgulation COmpliance Model

The DAPRECO (Data Protection Regulation Compliance) project (Luxembourg) aimed at the creation of a knowledge base for formal compliance with the terms and provisions of the GDPR supported by the creation of a Data Protection Ontology.

> *...the ontology will constitute the knowledge base from which the concepts to annotate the workflow model are extracted. Such an approach can provide benefits for a number of stakeholders:*

[21] https://harshp.com/research/publications/035-representing-activities-processing-personal-data-consent-semweb-gdpr-compliance#sec:contributions:ontologies

[22] Pandit, H. J., & Lewis, D. (2017). Modelling Provenance for GDPR Compliance Using Linked Open Data Vocabularies. *In Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn)*. http://ceur-ws.org/Vol-1951/PrivOn2017_paper_6.pdf.
Pandit, H. J, O'Sullivan, D. & Lewis, D. (2018). GDPR-Driven Change Detection in Consent and Activity Metadata. In *Joint Proceedings of the 4th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW), the 2nd Workshop on Semantic Web Solutions for Large-Scale Biomedical Data Analytics (SeWeBMeDA), and the Workshop on Semantic Web of Things for Industry 4.0 (SWeTI) Co-Located with 15th European Semantic Web Conference (ESWC 2018)*, 5. http://ceur-ws.org/Vol-2112/mepdaw_paper_2.pdf.
Pandit, H. J., O'Sullivan, D., & Lewis, D. (2018). An Ontology Design Pattern for Describing Personal Data in Privacy Policies. In *Proceedings of the 9th Workshop on Ontology Design and Patterns (WOP 2018) Co-Located with 17th International Semantic Web Conference (ISWC 2018)*. http://ceur-ws.org/Vol-2195/pattern_paper_3.pdf.
Pandit, H. J., O'Sullivan, D., & Lewis, D. (2018). Towards Knowledge-Based Systems for GDPR Compliance. In *Proceedings of the Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics (CKGSemStats)*. http://ceur-ws.org/Vol-2317/article-09.pdf.
Pandit, H. J., O'Sullivan, D., & Lewis, D. (2018). Queryable Provenance Metadata for GDPR Compliance. In *Procedia Computer Science*, 137:262–68. *Proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria*. https://doi.org/10/gfdc6r.
Pandit, H. J., O'Sullivan, D., & Lewis, D. (2018). GDPR Data Interoperability Model. In *23rd EURAS Annual Standardisation Conference*. http://openscience.adaptcentre.ie/pb/EURAS2018/.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

– data controllers would have a clearer view of their duties with respect to data protection in the context of their business;
– the auditors would have a first-look model to assess the GDPR compliance;
– DPAs would have a structured approach to detect potential violations.

This ontology reused core concepts from LKIF Core (Hoekstra et al. 2007[23]) to represent rights, rules, legal and natural persons.

In a later description, the so called DAPRECO model is described to consist of three distinct and connected components:

1. a machine-readable version of the legal text (Akoma Ntoso)
2. a semantic context of the law, i.e., a legal ontology (PrOnto)
3. a machine-readable and machine-processable representation of the logic of the legal provision (DAPRECO Knowledge Base of RIO Logic Formulæ)

While Akoma Ntoso is used to represent the GDPR text in machine-readable format and the representation of the logic of the legal provisions is done by the DAPRECO Knowledge base, the Privacy Ontology (PrOntO, see below) is the ontology being used by the model. This ontology not only includes the concepts used in data protection, as before, but also privacy-related concepts and it is not focused solely on GDPR.
Finally, the modelling of the GDPR concepts was carried out manually by a legal expert.[24]

PrOnto (Privacy Ontology)

As read in the previous section, the P**rOnto (Privacy Ontology)** provides concepts regarding legal privacy compliance associated with data types and documents, agents and roles, processing purposes, legal bases, processing operations, and deontic operations for modelling rights and duties.

It has a modular architecture, reuses existing vocabularies like the LKIF Core, ALLOT, and PWO ontologies, and was developed using the MeLOn (Methodology for building Legal Ontologies) during the participation in the European Horizon 2020 MIREL project.
The PrOnto ontology is not available online for reuse, we include it as a reference as it has been used in various projects.[25]

[23] Hoekstra, R., Breuker, J., Di Bello, M., Boer, A. et al.(2007). The LKIF Core Ontology of Basic Legal Concepts. *LOAIT 321*, 43-63.

[24] Bartolini, C., Muthuri, R., & Cristiana, S. (2015). Using Ontologies to Model Data Protection Requirements in Workflows. In *JSAI International Symposium on Artificial Intelligence*. http://orbilu.uni.lu/handle/10993/22383.

[25] Palmirani, M., & Governatori, G. (2018). Modelling legal knowledge for GDPR compliance checking. *Frontiers in Artificial Intelligence and Applications*, 313, 101-110. https://doi.org/10.3233/978-1-61499-935-5-101
Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., & Robaldo, L. (2018). Legal ontology for modelling GDPR concepts and norms. *Frontiers in Artificial Intelligence and Applications*, 313, 91-100. https://doi.org/10.3233/978-1-61499-935-5-91.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

There are other privacy-related ontologies developed from the legal perspective[26]:

- The **LegLOPD ontology** aimed at the preservation of privacy of users in location-based services. It modelled concepts from the Spanish data protection law.
- The **OntoPrivacy ontology** modelled the concepts of the Italian Personal Data Protection Code; a bottom-up approach was used as the lexicon was the basis to build the ontology.
- The application-oriented **Neurona ontologies** modelled the knowledge for the development of data protection compliance to offer reports regarding the correct application of security measures to data files containing personal data.

The Compliance Ontology (Information Model Ontology & Policy Model Ontology)

The Compliance Ontology was also developed as part of the European Union Horizon 2020 project (BPR4GDPR).

> *The **Compliance Ontology** is to provide the appropriate formalisms describing all important aspects related with compliance, focusing on fundamental entities, their categories, and their relations. In this context, the Compliance Ontology serves as the ontology of the GDPR universe, or, in more practical terms, a high-level codification of the GDPR into concepts that need to be taken into consideration by the BPR4GDPR policy framework,* as well as by the privacy-aware process re-engineering.

As part of this effort the Information Model Ontology was developed as the core of the Compliance Ontology, which covers the basic concepts, and the model in terms of entities and relations. While modelled in OWL format, the ontology has not yet been made available. The Policy Model Ontology was used to implement a rule-based framework on top of the Information Model Ontology.

Palmirani, M., Martoni, M., Rossi, A. Bartolini, C., & Robaldo, L. (2018). PrOnto: Privacy Ontology for Legal Compliance. In *Proceedings of the 18th European Conference on Digital Government ECDG 2018*, 10.
Palmirani, M., Martoni, M., Rossi, A. Bartolini, C., & Robaldo, L. (2018). PrOnto: Privacy Ontology for Legal Reasoning. In Andrea Kő $ Enrico Francesconi, *Electronic Government and the Information Systems Perspective* (pp. 139–52). Lecture Notes in Computer Science. Springer International Publishing.
Bartolini, C., Calabró, A., & Marchetti, E. (2019). Enhancing Business Process Modelling with Data Protection Compliance: An Ontology-Based Proposal. In *Proceedings of the 5th International Conference on Information Systems Security and Privacy* (pp.421-428). SCITEPRESS - Science and Technology Publications. https://doi.org/10/gf3czj.

[26] Casanovas, P., Rodríguez-Doncel, V., Cristiana, S. et al. (2016). A European framework for regulating data and metadata markets. *CEUR workshop proceedings*, 1750. http://ceur-ws.org/Vol-1750/paper-04.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

> *Rules are implemented as instances of the Permissions, Prohibitions and Obligations classes (sub-classes of the Rules class), whereas actions are implemented as Actions class instances, with ⟨a, op, res, org⟩ being reproduced by means of the corresponding object properties.*

While it is not available directly for reuse it does have an extensive specification.[27]

PROVO extension: GDRP Data Provenance Model

Ujcich, Bates, and Sanders identified six rights and obligations in the GDPR text and created relevant GDPR subclasses of PROV-DM agents, activities, and entities and encoding GDPR semantics into PROV-DM relations.[28]

> *The PROV Ontology (PROV-O) expresses PROV Data Model using the OWL2 Web Ontology Language (OWL2). It provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. It can also be specialized to create new classes and properties to model provenance information for different applications and domains.*[29]

Later, Campagna et al (2020) 30 Investigated the limitations of this extension and proposed some additional extensions to Ujcich et al work. While they proposed additional extensions to the model, they also highlighted the fact that they were conducting high-level efforts (analysis of GDPR text) void of practical approaches and the need to create a community around these tasks.

> *The path to helping organizations achieve true compliance should consider the high-level criticism aimed at improving a provenance model conjugated with practical measures that involve multidisciplinary and practical issues. […] Finally, although the GDPR will always require some human activity, we believe that collaborative efforts can result in a mature solution that will minimize bureaucratic tasks and maximize the transparency of processes.*

GDPR-related Ontologies for Specific Technical Domains

[27] Lioudakis, G., & Cascone, D. (2019). D3.1 Compliance Ontology. BPR4GDPR.
Dellas, N. (2019. D2.3 Initial Specification of BPR4GDPR Architecture. BPR4GDPR.

[28] Ujcich, B. E., Bates, A., & Sanders, W. H.(2018). A Provenance Model for the European Union General Data Protection Regulation. In K. Belhajjame, A. Gehani, & P. Alper (Eds.), *Provenance and Annotation of Data and Processes*,11017, 45-57. Springer International Publishing. https://doi.org/10.1007/978-3-319-98379-0_4.

[29] https://www.w3.org/TR/prov-o/
Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013). *PROV-O: The PROV Ontology*. https://www.w3.org/TR/prov-o/

[30] Campagna, D. P., Da Silva, A. S., Braganholo, V. (2020). Achieving GDPR Compliance through Provenance: An Extended Model. *In*: *Simpósio Brasileiro de Banco de Dados* (SBBD),35, 13-24. Sociedade Brasileira de Computação. https://doi.org/10.5753/sbbd.2020.13621.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

We briefly include in this review a few ontologies developed for the incorporation of GDPR knowledge in specific technical domains, such as sensor networks, blockchain, etc.

Fiesta-Priv

The **Fiesta-Priv ontology** extends the Semantic Sensor Network Ontology (SSN) with some GDPR requirements.

> *Additionally, the proposed ontology does not intend to cover all GDPR requirements with respect to the full data cycle (data gathering, sharing, processing, retention, deletion, etc.). Our main focus is on the data gathering and sharing aspects, providing entities that cover the phases of when (or not) the data of a user should be gathered and to which other users/applications these data should be shared[31].*

Fiesta-Priv was developed in the context of the FIESTA Project, and focuses on reuse of existing ontologies, and the representation of ISO principles combined with the requirements of GDPR.

> *With the introduction of GDPR, IoT systems and experimentation platforms that federate data from different deployments, testbeds and data providers must be privacy enabled. The wide adoption of IoT applications in many scenarios from smart cities to Industry 4.0 has raised concerns with respect to the privacy of users' data that are gathered from IoT devices. Many experimental smart city applications are also using crowdsourcing data. Inspired by the GDPR requirements, we propose an IoT ontology built using available standards that enhances privacy, enables semantic interoperability between IoT deployments and supports the development of privacy-preserving experimental IoT applications.[32]*

BIoT Ontology

Proposed in a recent Thesis dissertation, the Blockchain-based ontology for Internet of Things Security (BIoT) *to ensure an adequate level of security.*[33]

> *...the ontology provided insight into security properties to monitor vulnerabilities in the IoT ecosystem and blockchain network structure, thereby ensuring data integrity, confidentiality, and privacy.*

Experts participated in the knowledge acquisition process of this ontology.

[31] Agarwal, R., Elsaleh, T., & Tragos, E. (2020). GDPR-inspired IoT Ontology enabling Semantic Interoperability, Federation of Deployments and Privacy-Preserving Applications. https://arxiv.org/pdf/2012.10314.pdf

[32] Project github: https://github.com/ragarwa2/fiesta-priv

[33] Mendonça, S. F. T. O. (2019). *A blockchain-based ontology for the internet of things security.* [Thesis]. Universidade Federal de Pernambuco. https://repositorio.ufpe.br/handle/123456789/35813

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

> *The main requirements to participate in the study were: i) have a minimum of three years of experience (theoretical or practical) in the Internet of Things, Embedded Systems; ii) know Blockchain mechanism or technologies*

Recently, more PhD Thesis take an interest on privacy protection for both IoT and blockchain. For example:

- Moreira da Costa, T. (2017) *OPP_IoT An ontology-based privacy preservation approach for the Internet of Things*.[Thesis] Université Grenoble Alpes, https://tel.archives-ouvertes.fr/tel-01681206/file/MOREIRA_DA_COSTA_2017_archivage.pdf

- Arruda, M. F. (2019). *Um modelo ontológico e um serviço de gerenciamento de dados de apoio à privacidade na Internet das Coisas.* [Dissertação Mestrado em Ciência da Computação]. Universidade Federal de Goiás, Goiânia, http://repositorio.bc.ufg.br/tede/handle/tede/9323

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

APPENDIX C

Table 1. Blockchain and privacy CNIL comparison

Privacy problems in Blockchain	Legal Risk	CNIL Recommendation	OntoRopa
Identification of Data Controller	All participants may be qualified as data controllers when the processing is related to a professional or commercial activity (i) as natural persons, (ii) as legal persons, (iii) as "joint controllers".	To identify the data controller in advance (a representative or a legal person).	
Identification of Data Processors	In blockchain, smart contract developers and miners are deemed to be processors under GDPR	Processors and miners should establish a contract with the participant acting as data controller which specifies each party's obligations	
Identify the reasons to use blockchain solutions over other possible instruments	Not to comply with all requirements and safeguards set by GDPR	Favouring other solutions that allow for full compliance with the GDPR.	
Consider the requirements that affect data transfers outside the EU	The requirement for appropriate safeguards for transfers outside the EU, such as binding corporate rules or standard contractual clauses, are entirely applicable to permissioned blockchains	Permissioned blockchains should be favoured as they allow a better control over personal data governance.	
Carefully choose the format under which the data will be registered	In blockchain, the data registered on a blockchain cannot be technically altered or deleted once a block in which a transaction is recorded has been accepted by the	Some technical solutions should be examined by stakeholders in order to solve this issue.	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

	majority of participants.		
Identifiers of participants and miners	The architecture of blockchains means that these identifiers — alphanumeric characters which constitute the public key linked to a private key, known only by the participant— are always visible.	This data cannot be further minimised and that their retention periods are, by essence, in line with the blockchain's duration of existence.	
Additional data (or payload) stored on the blockchain containing personal data related to other individuals	The GDPR principle of data protection by design requires the data controller to choose the format with the least impact on individuals' rights and freedoms.	The CNIL considers that personal data should be registered on the blockchain preferably in the form of a commitment34, or alternatively in the form of a hash generated using a hash function with a key, or, at least, in the form of an encryption ensuring a high level of confidentiality.	
To ensure the effective exercise of rights	The GDPR was designed to give individuals back their control over personal information. The right to erasure, the right to rectification and the right to object to a blockchain are difficult to apply in blockchain.	The format chosen to register the data on a blockchain can also facilitate the exercise of individual rights.	
Compatibility of rights	The GDPR rights of information, of access and of portability are not problematic.	The data controller must provide concise information that is easily accessible and formulated in clear terms to the data subject before	

[34] A "commitment" is a cryptographic mechanism that allows one to "freeze" data in such a way that it is both possible - with additional information - to prove what has been frozen and impossible to find or recognise such data by using this sole "commit".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

		submitting personal data to miners for validation.	
Incompatible rights	It is technically impossible to grant the request for erasure made by a data subject when data is registered on a blockchain	However, when the data recorded on the blockchain is a commitment, a hash generated by a keyed- hash function or a ciphertext obtained through "state of the art" algorithms and keys, the data controller can move closer to the effects of data erasure using commitment schemes[35] and deletion of the keyed hash function's secret key.	
Security requirements	The different properties of a blockchain (transparency, decentralisation, tamper-proof and disintermediation) mainly rely on two factors: the number of participants and miners, and on a set of cryptological mechanisms.	For permissioned blockchains, the CNIL recommends: (i) Carrying out an evaluation of the minimal number of miners which would ensure the absence of a coalition that could control over 50% of powers over the chain; (ii) setting out technical and organisational procedures to limit the impact of a potential algorithm failure (including an emergency plan); (iii) the governance of changes to the software used to create transactions and to mine should be documented	

[35] "When a commitment scheme is perfectly hiding, deleting the witness (i.e. the element that allows to verify that a given value is committed in a given commit) and the value committed is sufficient to render the commitment anonymous in such a way that it can no longer be considered personal data".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

		(ensuring an alignment between planned permissions and practical application).	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338