

Privacy Compliance with Ontologies and Blockchain: The OntoROPA Project

M. Mercedes Martínez-González^{1,2,*,†}, Pompeu Casanovas^{2,3}, María-Luisa Alvite-Díez⁴, Núria Casellas², Amador Aparicio¹ and David Sanz¹

¹Universidad de Valladolid, Departamento de Informática, Campus Miguel Delibes, Valladolid, Spain

²Institute of Law and Technology - IDT, Universitat Autònoma de Barcelona, Bellaterra-Cerdanyola, Spain

³Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC), IIIA-IDT Associated Unit

⁴Universidad de León, Departamento de Biblioteconomía y Documentación, Campus de Vegazana, León, Spain

Abstract

OntoROPA (Ontological Records of Processing Activities) is a project to facilitate smart privacy legal compliance using technology capable of providing semantics, intelligence, and trust. OntoROPA deals with the creation and maintenance of a critical piece of legal compliance required by the General Data Protection Regulation (GDPR), the Records of Processing Activities (ROPA). OntoROPA's 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.

Keywords

Privacy, Compliance, General Data Protection Regulation (GDPR), Ontologies, Blockchain, Security, Trust

1. Introduction

Providing Records of Personal Data Processing Activities (ROPAs) is a mandate of the General Data Protection Regulation (GDPR), Regulation (EU) 2016/679, which rules ROPAs in Article 30. ROPAs are a critical piece of legal compliance required by the GDPR. These records are an instrument of legal compliance for private and public individuals and organizations that manage personal data. They provide an inventory of the data processing activities performed on private data. Keeping such records is an obligation for controllers and processors. ROPAs must contain a specific amount of information and they have to be kept in electronic form. Currently, most of these records are created at hand by the personnel in charge of this task—at best assisted by some legal compliance tool—, maintained in word documents or excel files, and made available to the public mostly in PDF or in their original formats.


Proceedings. Artificial Intelligence Governance Ethics and Law (AIGEL) Reviewed, Selected Papers, November 02 - December 19, 2022, Barcelona, Spain

✉ mercedes@infor.uva.es (M. M. Martínez-González); pompeu.casanovas@uab.cat (P. Casanovas); luisa.alvite@unileon.es (M. Alvite-Díez); ncasellas@nuriacasellas.com (N. Casellas); amador@infor.uva.es (A. Aparicio); david.sanz@uva.es (D. Sanz)

🌐 <http://www.infor.uva.es/~mercedes> (M. M. Martínez-González)

🆔 0000-0002-3151-0842 (M. M. Martínez-González); 0000-0002-0980-2371 (P. Casanovas); 0000-0003-1490-8936 (M. Alvite-Díez); 0000-0003-2546-9246 (A. Aparicio); 0000-0001-7879-411X (D. Sanz)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

While these formats are both human- and machine-readable, they are not suitable for semantic interoperability, automatic knowledge extraction, or automatic verification. Therefore, current ROPAs are not linked to related or similar documents, their content is not validated, and semantic interoperability is not enabled. On the contrary, citizens should be provided with tools able to extract the knowledge they keep, and show this knowledge to users in understandable manners, which is indeed a GDPR request. ROPAs should not be independent and isolated pieces of information. They should be reliable sources of relevant information, linked, and available for intelligent knowledge extraction. Technology can help to make it possible.

OntoROPA aims at the creation of a ROPA knowledge graph that will include not only the legal requirements, but also the practical knowledge from the community of privacy and data protection experts, including lawyers, legal advisors and scholars, data protection officers, and rulers who are proficient in the creation and manipulation of ROPAs.

The notion of practical knowledge is crucial because this entails an implicit professional knowledge that must be elicited and made explicit in the knowledge acquisition process. This kind of knowledge is also modelled, as it encompasses the professional selection and understanding of legal normative texts and provisions. This is not to be found in legal documents containing positive law because it belongs to their legal experience. This includes the interpretation of hard law, soft law, policies, and ethics.

2. Related work

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 data economy, compliance has also been bootstrapped, because if humans must be compliant, so must be 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 in 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” [1]. 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 [2] 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 Internet of Things—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 crisis, Fig. 2) and 2020-21 (because of the enforced implementation of GDPR) are the tipping points of the increase and growing interest of industry and researchers to find CbD and CtD solutions [3]. Hashmi, Governatori, Lam, and Wynn [4] 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 [5] is important here, because it deals with ethical principles, and ethical principles deal with Artificial Intelligence, and it will be even more connected in the future, according to the first draft of the next EU Artificial Intelligence Act¹.

2.2. Expert Knowledge as Methodology for ontology creation

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 [6] and Methontology [7] but include expert-centred and empirically oriented methods towards professional legal knowledge acquisition,

¹Recital n. 7 of the Draft reads: "In October 2020, the European Parliament adopted a number of resolutions related to artificial intelligence, including on ethics, liability, copyright, artificial intelligence in criminal matters, and artificial intelligence in education, culture and 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.

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.

2.3. Review of GDPR Ontologies

Since the 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 Data²) 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/EC³), many have pursued the encoding of their semantics for the development of smart data privacy compliant applications. We mainly mention 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. For extensive accounts of data protection related ontologies see [8] and [9]. 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. Existing ontologies are, namely: (i) SPECIAL usage policy language (OWL2); (ii) Data Privacy Vocabulary (RDF/OWL); (iii) Policy Log Vocabulary (RDF/OWL); (iv) DVP-GDPR (RDF/OWL); (v) GConsent (OWL2); (vi) GDPRov (OWL2); (vii) GPRtEX (RDF/OWL(SKOS)); (viii) Data Protection Ontology (OWL); (ix) PrOnto (Privacy Ontology) (OWL); (x) Compliance Ontology/Information Model Ontology/Policy Model Ontology (OWL); (xi) Fiesta-Priv ontology (OWL); and, (xii) BiOT (OWL).

2.4. Legal compliance

In 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 identify that "RegTech is about the digital tools that are necessary to master regulatory complexity" [10]. 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 [11]. 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.

²<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046>

³<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC>

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 Covid-19 pandemic, LawTech venture capital investments increased dramatically at the rate of 2.4 new start-ups per day [12]. The legal database hosted by CodeX, the Stanford Center for Legal Informatics keeps track of them.

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 [13]. However, turning norms from natural to formal languages combining NLP techniques and defeasible logic is a difficult task [14]. 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 [15]. This is a conceptual challenge, lately also called Rules as Code in e-government administrations [16].

3. The OntoROPA solution for privacy compliance

3.1. Approach

OntoROPA follows the general trend of the legal compliance market presented in section 2.4, but instead of interpreting directly the content of article 30 and article 33 of GDPR (mainly about the duties of controllers), an indirect way of approaching the subject was chosen:

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;

Table 1

Relation of OntoROPA goals and innovation with current problems of ROPAs

| Goal | Current issues | OntoROPA innovation |
|-----------------------------|--|--|
| ROPAs as knowledge graphs | No ontology collects knowledge from privacy compliance professionals | Professional ontology |
| Sharing, linking, reasoning | Isolated information cannot be used for knowledge inference | Community-based intelligence |
| Legal proof | No valid certification No proactiveness No content checks Privacy compliance in blockchain technology | Automatic certification Automatic proof Automatic verification Assisted ROPA creation |

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;
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.

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. Table 1 summarises its goals, which are aligned with the detected issues, and innovation.

3.2. Community

The target community of users starts with ROPA providers (ROPA controllers). The OntoROPA ecosystem will support more communities of ROPA users. For example, data protection supervisors can assess ROPAs. Figure 1 summarizes the flow of ROPAs within these communities in

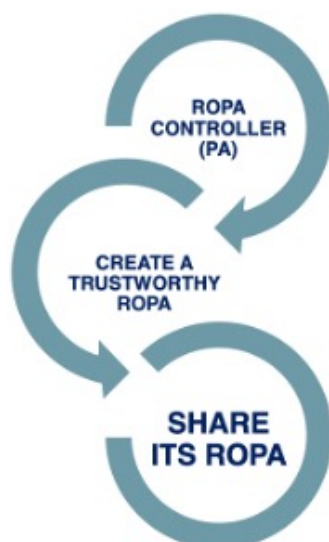


Figure 1: The OntoROPA community

OntoROPA. However, citizens are not able to assess ROPAs, but to read and query the information that ROPAs can provide to them about the way their personal data are treated and protected. A general solution, able to support different communities, requires a long-term project. The final solution entails the creation of a Law Tech legal web service to provide automated ROPAs to law firms, companies and administrations. This also entails the definition of a business model that fits into the niche of Data Protection and Privacy Services, as advanced by the European Digital Markets strategy.

3.3. Components and ontology

There are two main components in the OntoROPA project: (i) an OWL ontology that collects the expert knowledge from the target domain (ROPA community) and is the tool directing the inference processes that support validation and trustworthiness; (ii) and the software artifacts that process ROPAs.

OntoROPA proposes the development of a domain ontology formally expressed in OWL that will be offered as open data, reliable, reusable, and extensible. This professional ontology will support the creation and validation of ROPAs. Validation will be twofold: RDF validation for correctness and OWL validation for completeness. As already stated, the ROPA Ontology does not only include legal but also professional knowledge extracted from 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.

As a proof of concept, Figure 2 draws the ROPA RDF description which can be validated for correctness, and a preliminary ontology sample that demonstrates the reasoning capabilities for completeness of legal-compliance standard validation. The ontology creation process is based on metadata and competency questions.

```

@prefix ropa: <http://www.ontoropa.org/ropa#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://www.ontoropa.org/ropa-data#c4cc78ad-a91a-4ba2-a16d-cb1071e513c4>
a <http://www.ontoropa.org/ropa#RecordOfProcessingActivity> ;
ropa:hasController <http://www.ontoropa.org/ropa-data#1f949248-18ae-4fd5-be21-38ac2e364843> ;
ropa:hasRepresentative <http://www.ontoropa.org/ropa-data#735922d1-5f53-4dda-8aac-0e1df7b69bbb> ;
ropa:hasProcessingPurpose <http://www.ontoropa.org/ropa-voc/processing-purposes#purpose13> ;
ropa:hasDataSubjectCategory <http://www.ontoropa.org/ropa-voc/data-subject-categories#subject-category22> ;
ropa:hasPersonalDataCategory <http://www.ontoropa.org/ropa-voc/personal-data-categories#data-category4> .

<http://www.ontoropa.org/ropa-voc/processing-purposes#purpose13>
a ropa:ProcessingPurpose ;
skos:prefLabel "Tramitación de ayudas y subvenciones"@es ;
skos:definition "Tramitación de las ayudas y subvenciones gestionadas por la Dirección General de Competitividad de la Industria Agroalimentaria y de la Empresa Agraria."@es .

<http://www.ontoropa.org/ropa-voc/data-subject-categories#subject-category22>
a ropa:DataSubjectCategory ;
skos:prefLabel "Personas físicas"@es ;
skos:definition "Personas físicas, así como aquellas personas físicas que representen a las personas jurídicas, que tengan la condición de interesadas en las diferentes subvenciones y ayudas, que se gestionan por la Dirección General."@es .

<http://www.ontoropa.org/ropa-voc/personal-data-categories#data-category4>
a ropa:PersonalDataCategory ;
skos:prefLabel "Nombre"@es ;
skos:altLabel "Datos de identificación de las personas físicas: Nombre"@es .

ropa:PersonalDataCategory rdfs:subClassOf skos:Concept .
ropa:DataSubjectCategory rdfs:subClassOf skos:Concept .
ropa:ProcessingPurpose rdfs:subClassOf skos:Concept .

```

Figure 2: Example of RDF Description of a ROPA

3.4. Architecture diagram

OntoROPA uses a modular approach, where each module serves a specific functionality. This modular approach will facilitate OntoROPA resilience to changes in collaborators. For example, we can either take in charge the Identity module with the development of our own oracle, able to validate X509⁴ digital certificates in LDAP services or to use services provided by external providers and blockchain platforms.

A very important component of OntoROPA are data: ROPAs, ontologies, and data that helps to achieve the desired facilities, such as certificates and credentials used for identity verification. Figure 3 includes the data layer and software modules of OntoROPA. These data are critical for OntoROPA modules: they are inputs and outputs. More important, they determine the design of each module. This is a data-driven design. They can be described as follows:

1. **Identity**:: Legal compliance requires being able to link responsibilities and authorship to legal entities, real world entities. X.509 certificates will be used. Verification of these certificates requires to query LDAP directories.
2. **Linked RDF ROPAs**: 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

⁴X.509 certificates are digital certificates that use the widely accepted international X.509 public key infrastructure (PKI) standard to verify that a public key belongs to the hostname/domain, organization, or individual contained within the certificate.

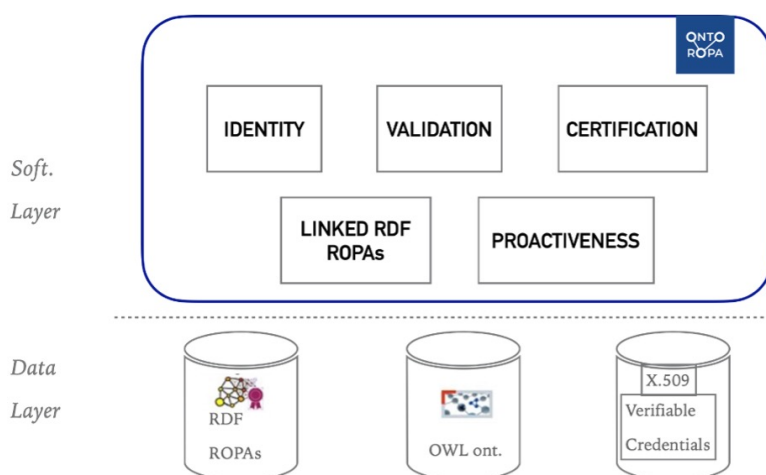


Figure 3: Main OntoROPA modules and data

Semantic Web standards provide the tools to represent, share and manage semantics in technical environments. Storage will rely on the facilities provided by a solution able to store and manage RDF graphs on blockchain. If not possible, an external RDF store, e.g. AllegroGraph, may be needed.

3. **Validation:** ROPAs should comply with article 30 of the GDPR2 and with the non-written rules of use that the community of experts, ROPA controllers, follow when they create them. This knowledge is collected in the OntoROPA ontology. The validation will be done against the ontology, using the inference capabilities associated to OWL rules and inference. We would like this validation to be a secure process, not subject to injections. WebProtégé can be used to reach this objective. The proof of validation will be taken in charge off-chain by OntoROPA, with its own signature and certificates.
4. **Certification:** ROPAs' origins and provenance should be certified. As for the results of validation, it depends on the viability of secure executions. If the process can be secured in a blockchain TEE, the blockchain enclave signature should reinforce OntoROPA's signature.
5. **Proactiveness:** The date in which a ROPA is available matters from a legal perspective. The immutability properties of blockchain platforms will support this. The transaction associated to the publication will provide proof of proactiveness.

4. A use case scenario

4.1. New ROPA flowchart

We introduce scenarios as the process of generating a ROPA in the context of the OntoROPA project. For instance, a person responsible of its creation and maintenance in an organization — the data privacy officer in a given university— needs to create and publish a ROPA to describe

personal data treatments in her university. These are the requirements: (i) using standard vocabularies, (ii) making sure that her ROPA includes the right information as required by article 30 of GDPR, and (iii) once this is achieved, publishing it and making it available to other ROPA providers, data protection supervisors, and the public in general (this is mandatory for Public Administrations). Moreover, she wants to be able to give evidence on the publication date if the data protection supervisory authority (in Spain, the AEPD; in France, the CNIL, etc.) launches an inspection after critical situations such as data breaches⁵. The data privacy officer may use the application providing the way of creating a ROPA.

There are two main possibilities: (i) Import ROPA: A ROPA is already available as a pdf or excel sheet. This ROPA is imported; (ii) New ROPA: A ROPA is created from the beginning.

We will elaborate on the second one: A ROPA provider wants to create a new RDF ROPA to describe the activities dealing with personal data. Figure 4 shows an overview of the process flow:

1. The first step is to create the RDF file describing the ROPA.
2. The second step is to validate the ROPA and check that it contains the right information as requested by the GDPR.
3. Once it is ready for publication, its quality is certified.
4. The certified ROPA is published.

Figures 4 and 5 show the flowchart and the sequence diagram of this use case. If the ROPA already exists, for example a ROPA that is already available as an excel file, it will be imported and an equivalent RDF description obtained. This RDF description is enriched with semantic metadata according to the ontology, which provides the semantic layer that the excel sheet is unable to provide. If the ROPA must be created from scratch, the ontology will be used to guide the creation and to verify the correctness of the new description. Once this is done, the RDF description is ready to be certified as a valid description. This is made in a Trusted Execution Environment for security reasons. If valid, a signature will provide evidence of it. The valid ROPA (ROPA + signatures) is ready to be published, which is done by uploading it to the blockchain. The blockchain platform is used to provide trust. Trust is a way of guaranteeing the date the ROPA was published, that is, that it has not been altered after some data breach or any other incident, or after the request issued by the data protection supervisor.

4.2. Functional requirements

OntoROPA functional requirements are summarised in Table 2. The main functionalities are creating, editing, and deleting a ROPA. But there are also some additional functionalities derived from the goals of providing legal validity to ROPAs: validation, signing, certification. Users who are authorized to create, edit, and publish ROPAs should be able to identify themselves.

⁵The GDPR sets the obligation of keeping available ROPAs for inspections if required by the data protection supervisor authority. Moreover, it introduces the concept of proactiveness, which means that (i) privacy by default and by design have been applied from the very beginning, (ii) that the security measures have been implemented, and (iii) the information about personal data activities is available. ROPAs are the records collecting this information. Therefore, ROPAs must be available while personal data treatment develops. Some data protection experts are concerned about the possibility that ROPAs are generated after the supervisor's request.

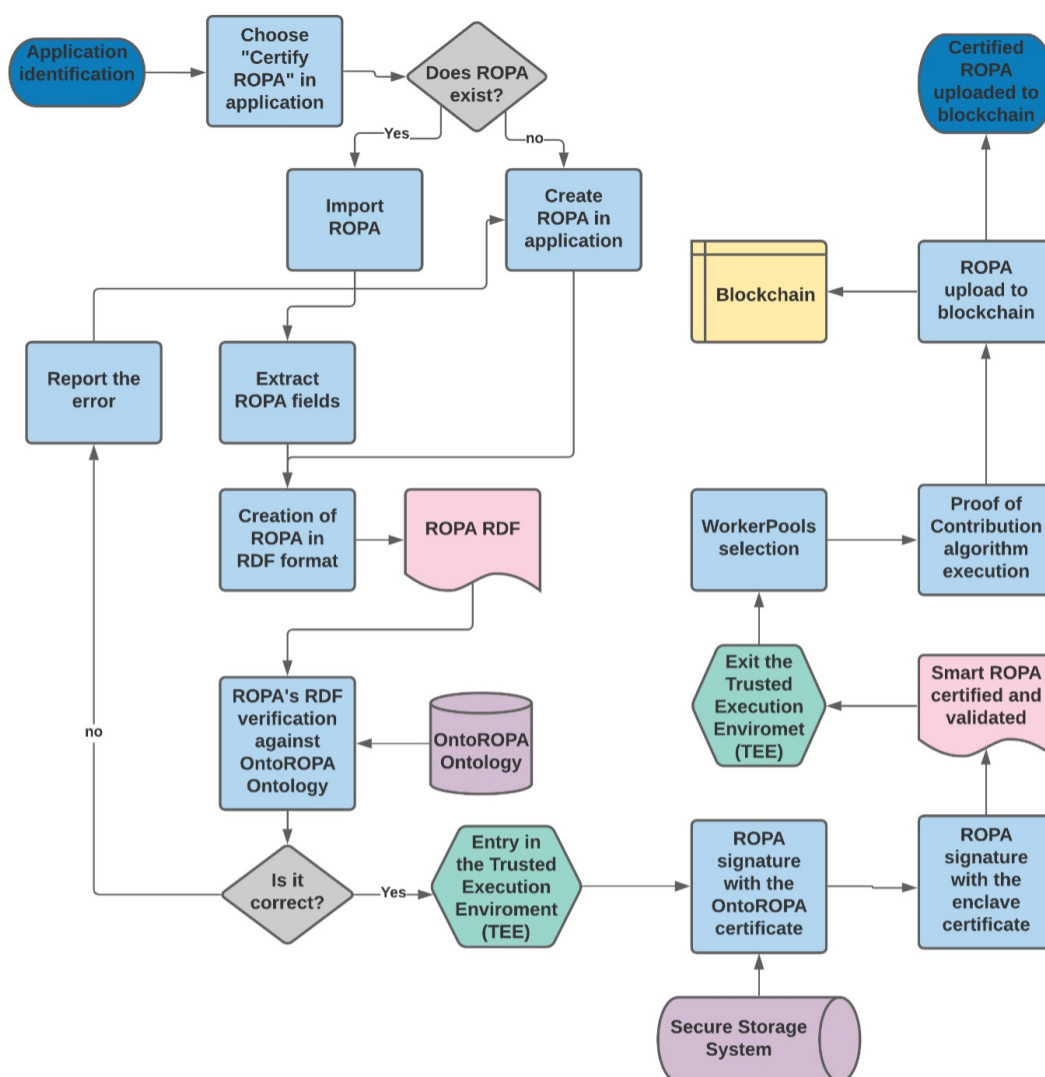


Figure 4: Flowchart for the New ROPA use case

5. Legal values

5.1. Return on Investment in Legal Compliance

Legal Compliance is essential in organizations to ensure compliance with their Codes of Conduct, as consumers demand products and services provided with "ethical and sustainable" behaviours. They often access social networks to publicly denounce those companies that do not meet their commitments, resulting in serious reputational damage and significant sales drops. Non-compliance with these obligations is punished with a range of sanctions ranging from heavy

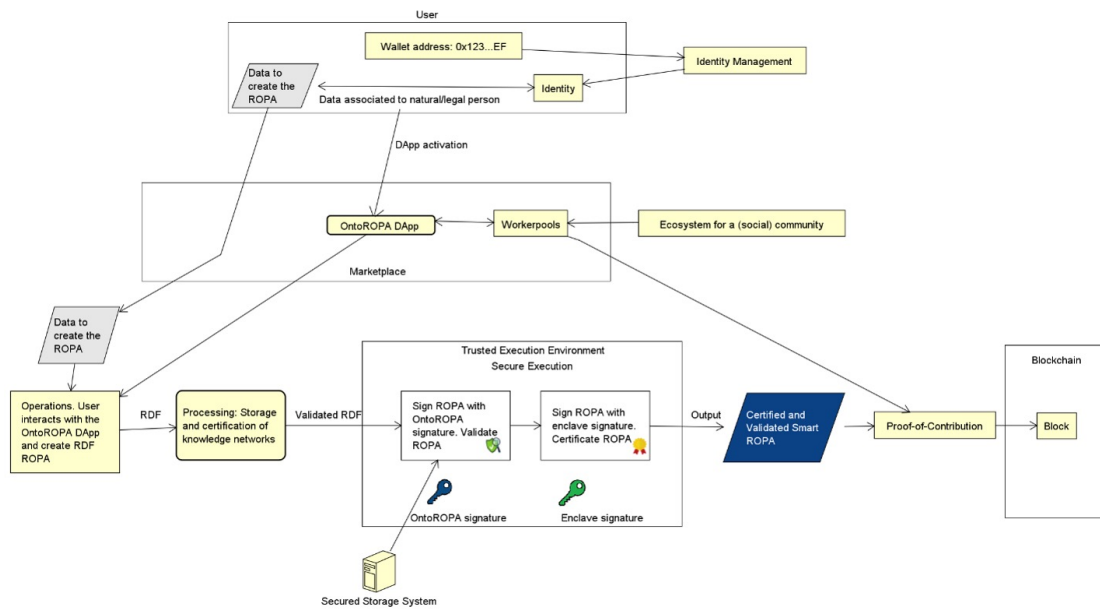


Figure 5: Data flow for a new ROPA use case

Table 2

Functional requirements in OntoROPA

| ID | Name | Description |
|-----|---------------|--|
| FR1 | Identity | Only users able to identify themselves as members of the ROPA enabled community will be authorized to create ROPAs |
| FR2 | Create ROPA | Create a new ROPA |
| FR3 | Edit ROPA | Modify a ROPA |
| FR4 | Delete ROPA | Erase a ROPA |
| FR5 | Validate ROPA | Check the correctness of a ROPA |
| FR6 | Sign ROPA | Sign ROPA with the digital signature of its creator |
| FR7 | Certify ROPA | OntoROPA certifies the validity of a ROPA with its signature |
| FR8 | Publish ROPA | A certified ROPA is published |

finances to professional disqualification or cessation of activity, as well as irreparable reputational damage.

There are three indicators that can help quantify a return on investment in legal compliance: (i) increased competence and efficiency within the organization; (ii) savings by reducing legal risks and prevention of sanctions; (iii) generation of better business opportunities. With automated and standardized toolkits, it becomes easier to meet the requirements of ISO 27001⁶ and ENS (Esquema Nacional de Seguridad)⁷. ROPA controllers can benefit from having a standard tool to simplify the task of creating their own ROPAs, and the possibility to adapt/extend it to their

⁶<https://www.iso.org/standard/54534.html>

⁷<https://www.boe.es/eli/es/rd/2022/05/03/311>

own use cases.

5.2. Legal validity

Legal validity (i.e. ‘legality’) is not equivalent to computational or logical validity. ROPA validation refers to the accuracy, traceability and technical reproductivity of the process that has generated it. It can be reached through the ontology.

However, this is not turning ROPAS into valid processes with legal outcomes and effects. Automated legal validity should be carried out aligning: (i) the selection of relevant legal sources in a transparent, shareable, and acceptable way, according to the main legal doctrine, (ii) the normative interpretation process that is accepted by official bodies, such as Data Protection agencies, (iii) as a last resort, the normative interpretation process that is accepted by regional, national, and European judiciaries. There are a variety of normative and regulatory sources that should be considered.

To ease the process of handling them we have defined them into four legal different clusters: (i) Hard law (laid down by Parliaments and the Judiciary (this includes European Regulations, such as the GDPR, and the Directives that have been transposed into the national legal systems by the State members); (ii) soft law (such as international agreements and covenants, mandatory after mutual or collective agreements); (iii) policies (issued by European and national governments to developing, enforcing, and implementing Acts, Regulations, and case-based law sentences), (iv) ethical principles and values, as they have been discussed, proposed and accepted in specific sectors (such as the recent EU guidelines for Artificial Intelligence).

Besides legislation, it is worth noting that the legal value—i.e. legal validity—is created through a process that fosters legal security and social trust among all stakeholders in the market (including companies, corporations, administrations and citizens). Then, ISO standards and technical protocols (such as the W3C standards and recommendations) matter. As stated by EU recent strategies, better regulation principles involving Impact Assessments and citizens’ consultations, and the introduction of digital currencies as a basis for the EU digital market fosters the general use of specific policies and best practices that benefit from the experiences already gathered.

OntoROPA embraces the middle-out approach to AI governance set by the AI4People Report to the EU Parliament (November 2019)⁸. It can be defined as the middle-ground between top-down and bottom-up regulatory approaches, fostering co-regulation, co-responsibility and dialogue between rulers and the subjects of regulation [17]. Certified and validated ROPAs are followed by a proof of contribution and a smart contract linking users, controllers, and supervisors, in between blockchain and the community of users.

It is worth mentioning that law or its digital version, legal governance systems [18], do not constitute in OntoROPA a third layer on top of the data layer and the software layer defined above (section 2). There is no legal layer consisting mainly in documents that can be deemed ‘legal’. What it does exist instead is a dynamic set of normative systems, guidelines, values, policies, standards, and best practices that integrate a complex cognitive system embedded into human behaviour and (now) information systems.

⁸https://www.eismd.eu/wp-content/uploads/2019/11/AI4Peoples-Report-on-Good-AI-Governance_compressed.pdf

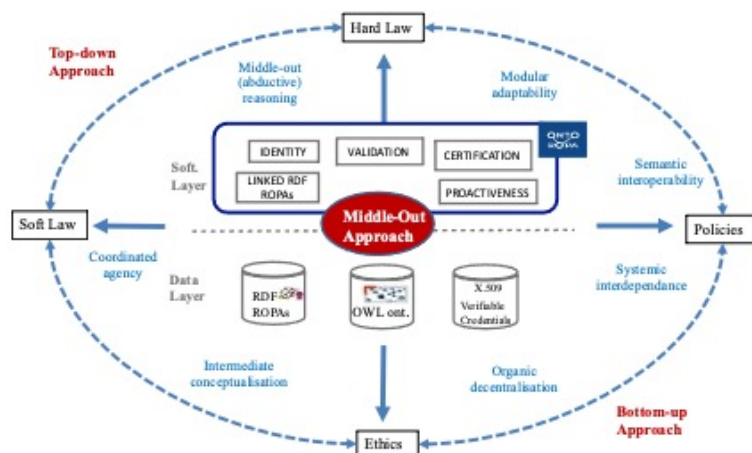


Figure 6: OntoROPA Legal Ecosystem

This dynamic set constitutes a dimension of human and artificial systems and interfaces. It pervades the software and the data layer from inside out. This is why a middle out approach can be the most appropriate to generate the legal ecosystem that is needed to validate ROPAs and ROPAs' computational management in both senses—technological and legal. There are two layers—software and data layer—and three dimensions—technological, social, and legal. The links between them occur stemming from the secured process to produce a certified and legally valid ROPA.

The OntoROPA legal ecosystem is generated by the set of technical requirements and social and legal conditions that are taken into account by controllers, supervisors, professional agents in the marketplace (legal web services, law firms and companies). Thus, the certification and validation processes involve the participation of all stakeholders. Again, technical requirements do not reflect per se the social and legal conditions. They are reached through (i) the mutual understanding of regulations, i.e. the shared agreement on the rights and duties set by the regulatory system (legislation, policies, best practices, and ethics), (ii) the mutual understanding of the position of all agents participating in the process, (iii) the mutual understanding of all necessary actions to be taken to make the final product 'legal'. This is where the legal validity of certification comes from. Certification and validation processes do not stand by their own: They are necessary components of the legal ecosystem generated through the coordination of all required elements, as shown in Figure 6.

The use of blockchain technologies has generated some controversies about its compatibility with GDPR requirements. Permissionless blockchains are distributed, decentralised peer-to-peer networks in which everyone can participate interacting with unknown counterparties, trusted or not [19]. 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 [20]. The study recommends 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”. These risks have been singled out for ROPAs’ implementation mechanisms [21]. In addition, as stated by the French Commission Nationale de l’Informatique et des Libertés (CNIL), there are legal risks that arise from this situation, i.e. uncertainty, blurred identification of liable stakeholders, and lack of a clear allocation of duties in case of multiple controllers. We do not have the solution yet for all the issues, but focusing on transactions and having in mind the certification process helps to sort them out.

6. Conclusions and Future Work

The OntoROPA project is law and data driven. ROPAs are deemed to be a critical piece of legal compliance from a social perspective, for they are the only available source of information accessible to non-technical people (including citizens, judges, rulers, law experts, data protection users, and supervisors). Thus, this fact makes them a critical piece for GDPR compliance for all stakeholders—providers, controllers, supervisors, and companies. This is a market niche. As a result, we figured out a legal governance system that facilitates a soft orchestration of hard law, soft law, ethics, and policies.

This also is work in progress. Some steps have been advanced in the implementation process, e.g. the transformation of ROPAs to designed semantic schemas has been tested, the first version of the ontology has been already built, and some tests with tools that can be used to provide trust have been carried out. There is still room for improvement. The design of blockchain tools, and the implementation of AI algorithms that will validate ROPAs against the ontology will be developed in the next future.

7. Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957338, *NGI OntoChain - Trusted, traceable and transparent ontological knowledge on blockchain*.

References

- [1] O. Akhigbe, D. Amyot, G. Richards, Information technology artifacts in the regulatory compliance of business processes: A meta-analysis, in: M. Benyoucef, M. Weiss, H. Mili (Eds.), *E-Technologies - 6th International Conference, MCETECH 2015, Montréal, QC, Canada, May 12-15, 2015, Proceedings*, volume 209 of *Lecture Notes in Business Information Processing*, Springer, 2015, pp. 89–104. URL: https://doi.org/10.1007/978-3-319-17957-5_6. doi:10.1007/978-3-319-17957-5_6.
- [2] P. Casanovas, J. González-Conejero, L. de Koker, Legal Compliance by Design (LCbD) and through Design (LCtD): Preliminary Survey, in: V. Rodríguez-Doncel, P. Casanovas, J. González-Conejero (Eds.), *Proceedings of the 1st Workshop on Technologies for Regulatory Compliance co-located with the 30th International Conference on Legal Knowl-*

- edge and Information Systems (JURIX 2017), Luxembourg, December 13, 2017, volume 2049 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017, pp. 33–49. URL: <https://ceur-ws.org/Vol-2049/05paper.pdf>.
- [3] O. Akhigbe, D. Amyot, G. Richards, A systematic literature mapping of goal and non-goal modelling methods for legal and regulatory compliance, *Requir. Eng.* 24 (2019) 459–481. URL: <https://doi.org/10.1007/s00766-018-0294-1>. doi:10.1007/s00766-018-0294-1.
- [4] M. Hashmi, G. Governatori, B. Lam, M. T. Wynn, Are we done with business process compliance: state of the art and challenges ahead, *Knowledge and Information Systems* 57 (2018) 79–133. doi:10.1007/s10115-017-1142-1.
- [5] L. Floridi, J. Cowsls, A Unified Framework of Five Principles for AI in Society, *Harvard Data Science Review* 1 (2019). doi:10.1162/99608f92.8cd550d1.
- [6] N. F. Noy, C. D. Hafner, Ontological Foundations for Experimental Science Knowledge Bases, *Appl. Artif. Intell.* 14 (2000) 565–618. URL: <https://doi.org/10.1080/08839510050076972>. doi:10.1080/08839510050076972.
- [7] M. Fernández-López, A. Gómez-Pérez, N. Juristo, METHONTOLOGY: From Ontological Art Towards Ontological Engineering, in: *AAAI Conference on Artificial Intelligence*, 1997, pp. 33–40. URL: <https://oa.upm.es/5484/>.
- [8] P. Ryan, H. J. Pandit, R. Brennan, A Common Semantic Model of the GDPR Register of Processing Activities, *CoRR abs/2102.00980* (2021). URL: <https://arxiv.org/abs/2102.00980>. arXiv:2102.00980.
- [9] B. Esteves, V. Rodríguez-Doncel, Analysis of ontologies and policy languages to represent information flows in GDPR, *Semantic Web* 1 (2022). doi:10.3233/SW-223009.
- [10] M. Nedelchev, Origin and Definition of RegTech, *Economics and Management* (2020). doi:10.37708/em.swu.v17i1.1.
- [11] A. D. Kristanto, A. A. Arman, Towards A Smart Regulatory Compliance, The Capabilities of RegTech and SupTech, *2022 International Conference on Information Technology Systems and Innovation (ICITSI) (2022)* 300–309. doi:10.1109/ICITSI56531.2022.9970801.
- [12] P. Casanovas, *Inteligencia Artificial y Derecho: La doble implosión de las profesiones y servicios jurídicos en la era digital*, Centro de Investigaciones Sociológicas (CIS), Madrid, 2022, pp. 83–114.
- [13] G. Boella, S. C. Tosatto, S. Ghanavati, J. Hulstijn, L. Humphreys, R. Muthuri, A. Rifaut, L. van der Torre, Integrating Legal-URN and Eunomos: Towards a Comprehensive Compliance Management Solution, in: *International Workshop on AI Approaches to the Complexity of Legal Systems, AICOL 2013, Lecture Notes in Computer Science*, volume 8929, Springer, Berlin, 2013, pp. 130–144. doi:10.1007/978-3-662-45960-7_10.
- [14] A. Z. Wyner, G. Governatori, A Study on Translating Regulatory Rules from Natural Language to Defeasible Logics, in: P. Fodor, D. Roman, D. Anicic, A. Z. Wyner, M. Palmirani, D. Sottara, F. Lévy (Eds.), *Joint Proceedings of the 7th International Rule Challenge, the Special Track on Human Language Technology and the 3rd RuleML Doctoral Consortium*, Seattle, USA, July 11 -13, 2013, volume 1004 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2013. URL: <https://ceur-ws.org/Vol-1004/paper16.pdf>.
- [15] L. Humphreys, G. Boella, L. van der Torre, L. Robaldo, L. D. Caro, S. Ghanavati, R. Muthuri, Populating legal ontologies using semantic role labeling, *Artif. Intell. Law* 29 (2021) 171–211. URL: <https://doi.org/10.1007/s10506-020-09271-3>. doi:10.1007/s10506-020-09271-3.

- [16] M. Waddington, Rules as Code, Law in Context. A Socio-legal Journal (2021). doi:10.26826/law-in-context.v37i1.134.
- [17] U. Pagallo, P. Casanovas, R. Madelin, The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data, *The Theory and Practice of Legislation* 7 (2019) 1 – 25. doi:10.1080/20508840.2019.1664543.
- [18] P. Casanovas, L. de Koker, M. Hashmi, Law, Socio-Legal Governance, the Internet of Things, and Industry 4.0: A Middle-Out/Inside-Out Approach, *J* (2022). doi:doi.org/10.3390/j5010005.
- [19] M. M. Martínez-González, P. Casanovas, M.-L. Alvite-Díez, I. Gutierrez, N. Casellas, *OntoROPA D1: State of the Art and Ambition*, Technical Report, European Commission, *OntoChain* (Grant 957338), 2021. doi:10.5281/zenodo.4930187.
- [20] E. Cirone, Blockchain and the General Data Protection Regulation: an irreconcilable regulatory approach?, *Queen Mary Law Journal* (2021). doi:10.26494/qmlj3939.
- [21] M. M. Martínez-González, P. Casanovas, M.-L. Alvite-Díez, N. Casellas, I. Gutierrez, *OntoROPA D2: Proposed Design Specification and Approach*, Technical Report, European Commission, *OntoChain* (Grant 957338), 2021. doi:10.5281/zenodo.4930887.