Towards Evidence-Based Conceptual Modeling for International Data Protection Requirements

1st Claudia Negri-Ribalta
IRiSC - SnT, Institute of Advanced Studies
University of Luxembourg
Luxembourg, Luxembourg
0009-0003-8480-5788

2nd Rene Noel
Escuela de Ingeniería Informática
Universidad de Valparaíso
Valparaíso, Chile
0000-0002-3652-4645

3rd Anastasia Sergeeva
IRiSC - SnT
University of Luxembourg
Luxembourg, Luxembourg
anastasia.sergeeva@uni.lu

4th Gabriele Lenzini
IRiSC - SnT
University of Luxembourg
Luxembourg, Luxembourg
0000-0001-8229-3270

***Abstract*—Data protection regulations worldwide impose various regulatory requirements on organizations, some overlapping and some differing. Identifying and tracking these requirements is vital for transborder data flows and compliance. Data Protection Impact Assessments (DPIAs) help translate regulations into software specifications and organizational policies, but they often use vague legal language, leading to misunderstandings.**

Conceptual modeling may support a shared understanding of the domain. Ontologies and modeling methods could help bridge the understanding gap among professionals with different backgrounds in data protection, particularly in transnational realities. Developing these tools requires theoretical knowledge and input from legal practitioners. By identifying common principles and requirements across regulations, practitioners can identify specifications requiring attention for transborder data flows. OBI-PIA aims to tackle this through interdisciplinary research, proposing a regulatory data protection ontology and conceptual modeling method to guide the DPIAs discussion process.

This paper presents a work-in-progress (WiP) based on interviews with legal practitioners worldwide. Preliminary results suggest that most regulations promote the OECD privacy principles, and specific requirements such as consent and the conceptualization of personal data. Inspired by the international relations literature, we propose categorizing regulatory data protection requirements into two groups: first-level (common requirements) and second-level (national, different) requirements as first step to star discussing DPIAs in transborder personal data flows. OBI-PIA should help practitioners identify requirements from each level, and discuss in interdiscplinary groups about compliance.

***Index Terms*—data protection, requirements, compliance, privacy**

I. Introduction and problem statements

Data protection regulations aim to protect the privacy (among others) aspect of personal data [1]. As a result, data protection regulations impose different types of requirements into information systems (IS), affecting the Software Development Lifecycle (SDLC) and data transfer across jurisdictions [2]. These regulations are often written in a general, unspecific, verbose, legal language that allows for multiple interpretations in different scenarios across times [3]. This situation sharply contradicts what IS requirements need to be: precise, well-defined, and measurable.

Despite ongoing efforts from the requirements engineering (RE) community in proposing data protection requirements (DPRs) artifacts, repeated fines from European and global regulators indicate a persistent failure to properly elicit and integrate these into IS. Multiple reasons may explain this, such as the difficulties in translating regulation into specification [2], [4]; different mental models of stakeholders [5]; beliefs that DPRs can be satisfied with security specifications [5]; or communication difficulties between lawyers and other stakeholders [6]. Having proper communicative artifacts that can help stakeholders conceptualize and discuss these requirements is useful.

In an increasingly globalized society, with continuous personal data flows across jurisdictions, correctly identifying, interpreting, and implementing DPRs is critical. Organizations must determine the legality of transborder personal data flows and resolve potential regulatory requirements conflicts to prevent costly IS redesigns [7]. Transborder personal data flows[1] are the movement of personal data across jurisdictions for different purposes, such as data storage, business or communication. Therefore, organizations must identify if they fall under the scope of data protection regulations and whether they offer services to residents or have processing activities within national borders, among others. The relevance of DPRs is critical for the interoperability of systems supporting transborder personal data flows, and organizations aiming for international operations may need to answer multiple questions [7], [9].

For example, an EU-based organization providing services

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions grant agreement 101081455 and by grant ANID/FONDECYT/INICIACION/11251489.

[1]The OECD [8] defines it as 'movements of personal data across national borders'.

to Brazilian citizens and using Californian-based cloud services from the USA will need to check and comply with at least three or more regulatory regimes. Do these regulations conflict? Do the regulations of the countries involved reflect a common conceptualization of what consent means? Can Brazilians' personal data be transferred or processed in an EU state and California? These may be some of an organization's many questions to answer, showing the complexity of aligning regulatory requirements across multiple jurisdictions, a concern recognized by the OECD [7], [10].

Software engineers have problems dealing with national regulations and communicating in their languages with lawyers. We can expect these difficulties to intensify across cultures and international regulatory regimes. Furthermore, DPR expectations vary across jurisdictions, which can increase communication difficulties [7], [10], [11].

A critical factor in transborder personal data flow is that legal experts must interpret regulations. The nuance of legal verbose is key for managing this international data flow. Lawyers' domain vocabulary is influenced by their country regulations, which may hinder common conceptualizations (of DPRs). Thus, identifying common and differing conceptualizations is vital to avoid costly reworks, fines, and reputation damage. Legal uncertainty, potential incompatibility among requirements, and the resources required to identify and address them represent key challenges faced by organizations [10].

This research preview deals with the design problem of how to use conceptual modeling techniques to help organizations to develop DPRs for international data privacy requirements. To address this challenge, we have started developing the OBI-PIA method. The OBI-PIA method is an initiative for developing a conceptual modeling method for the elicitation, analysis, specification and validation of DPR, founded on empirical evidence, that is currently in its early stages of development. The OBI-PIA method integrates computer science, law, and political science, and aims to help interdisciplinary teams at an international level involved in IS development to facilitate communication and understanding of regulatory DPRs, independent of the jurisdiction. Therefore, organizations can prioritize requirements that are different across regulations by identifying common DPRs across regulations, reducing compliance complexity and system rework costs.

Our research plan has been divided into three phases: (1) identification of common international regulatory DPRs; (2) conceptualization of these requirements; and (3) design of an evidence-based conceptual modeling method for international DPRs. The evidence-based approach comes from the research method for the first element: 80 to 120 interviews with legal experts from 28 countries (G20 countries, plus 7 randomly selected invitees and the host country of the research), serving as bases for designing conceptual artifacts. This research preview details the research definition and the preliminary results from 45 interviews. Hence, our contribution is to address data protection compliance concerns explicitly stated by law experts.

The remainder of the paper is organized as follows: Section II presents related work and previous research. Section III-C introduces the proposed research, including the main motivation, the research design and an overview of the OBI-PIA modeling method. Section IV presents some preliminary results of the ongoing research and examines possible threats, limitations and future work. Finally, Section VI summarizes our research proposal.

II. Background

A. Data protection worldwide

As privacy has various definitions [12], jurisdictions have enacted different regulations regarding processing personal data [1]. Similarly to privacy, data protection has been considered a human right, such as in the European Union Charter of Fundamental Rights [13]. Given the growing importance of data protection and its distinction from privacy, the legal instruments seeking to regulate it focus on managing personal data lifecycle beyond the privacy realm. Elements such as transborder data flows, business processes, fines, and definitions of institutions are imposed by data protection regulations that are not part of the conceptualization of privacy [1]. Consequently, organizations must adapt their IS to fulfill these constraints for compliance purposes, which if not, can lead to fines and reputation damages.

Expanding internationally poses challenges for organizations due to varying national regulations. A known example from the European perspective has been the *'Schrems saga'*, where the personal data transfer agreements between the USA and the EU were questioned (which led to the invalidation of the EU-US Privacy Shield and the new Data Privacy Framework) [14]. Because regulations are generally written at national levels, these requirements may drastically vary between jurisdictions, even contradict each other [10]. While the GDPR has become the *ipso facto* international standard on data protection according to some expert[2], it remains to be understood how all regulations have been affected by the Brussels effect [15] and authors disagreeing with the *Brussels effect*. From an RE perspective, research has primarily focused on regulations from Europe, Canada, China, the United States, and Brazil.

Consequently, when dealing with personal data flows from one regulation to another — a process usually called transborder data flows — failing to timely identify common, diverging, and conflicting requirements can lead to costly IS reworks, fines, and others. International organizations, such as the OECD and the APEC, have extensively worked on the subject [16]. The former has worked on the subject since the 1980s [17], and the latter has also developed its own principles, including the Cross-Border-Privacy Rules (CBPRs) [16]. Transborder personal data flows can boost the economy, market, and cooperation [18]; however, current analysis shows that stakeholders express concerns over the uncertainty of legal regimes (and their requirements), the incompatibility between them, and the times and resources demanded for transborder personal data flows [7], [10]. These concerns are

[2]Due to space constraints, we cannot discuss the historical reasons for this situation, of what was been recognized as the Brussels effect.

not, having been expressed several decades ago [19]; what appears novel is how it has extended to others process.As such, artifacts can help organizations with these concerns; there are some ongoing efforts in creating artifacts that can help with this transborder data flows, such as documentation, identification of the mechanism for transfer, legality, and monitoring, such as user-held data models [14].

B. Related work - Data protection requirements in the software development lifecycle

From a requirements engineering (RE) perspective, the literature agrees that requirements can originate from different sources, including regulations and laws [2]. However, software engineers seem to have difficulties translating regulatory DPRs, as they note they are too 'complex' [5]. Indeed, understanding regulatory DPRs requires specific expertise in legal prose and knowledge [3]. Grounding data protection can help have a shared abstraction of the phenomenon [5]. Multiple conceptual modeling artifacts have been proposed to work with regulatory data protection and privacy requirements [20]–[22]. Due to space constraints, we cannot fully describe all proposals on the subject.

Among them, ontologies can be essential in formalizing concepts, entities, and their relationships [23]. In RE particularly, they '... convert implicit shared understanding into explicit shared understanding', reducing misunderstandings [24]. While several ontologies exist for privacy requirements [22], it is crucial to recognize that privacy differs from data protection and has limitations [1]. For example, in [25], the PriS conceptual model indicates that data protection and security are subgoals of the privacy goal. Such ontological commitment can be debated from a regulatory perspective, as privacy is not the same as data protection [1], and equating the two may raise conflicting expectations between stakeholders [5]. From a purely data protection perspective, most ontological work focuses on particular and limited areas of regulations [26]. As [26] concludes, most ontologies partially model the GDPR, focus on specific aspects, and are outdated.

Some initiatives have proposed conceptual modeling languages and methods for addressing GDPR requirements, that is to say, DPRs. From a goal-oriented paradigm, the STS-ml framework has been extended for the GDPR [27]. While helpful for interdisciplinary [6], it only partially models the GDPR. BPMN and UML models have also been suggested for the GDPR by [28] and [29] accordingly,but as the authors indicate, it is also limited, and more work is required. These artifacts share the same limitation as the ontologies: they are limited to some specific jurisdiction requirements.

III. OBI-PIA:An Evidence-Based Conceptual Modeling Method for International Data Protection Requirements

A. The need for common conceptualizations

Organizations may have to comply with different jurisdictions because they fall under their territorial scope. Compliance may be due to territorial location, services provided, or personal data flows that fall under regulatory regimes. Taking the same scenario as the introduction – an EU-based organization providing services to Brazilian citizens, using Californian-based cloud services — exemplifies the multiple challenges that may surface: workers within the organization speak different natural languages, have different cultures (and thus DPR expectations [30]), ways of working and need to comply with different regulations. At first glance, in this scenario, the organization must verify compliance in at least three jurisdictions: Brazil, the USA, and the EU country (for example, France) where it is based.

In this context, requirement analysis becomes pivotal when dealing within an international context, where personal data flows may exist between jurisdictions. Organizations must assess the risks of such processes and identify common and divergent requirements across regulations. Can EU-based companies store personal data in a cloud in California? What are the rights of data subjects in Brazil and France? Are they the same? Failing to identify conflicting requirements can lead to unsatisfactory compliance levels, IS re-works, or fines. Therefore, discussing with stakeholders of different expertise, cultures, and languages (both technical and natural) is critical for transborder data flows.

However, as discussed in Section II, practitioners from different areas, such as law and engineering, have different mental models and vocabularies regarding DPRs [5]. Although implicit understanding [24] can help with the communicative process, transcultural shocks hurt this process and may take time to build. Additionally, when transcultural teams talk about DPRs, they may not refer to the same thing. '*The right to be forgotten*' may be conceptualized differently in Brazil, France, and California [7]. This different conceptualization, if not made explicit, can lead to misunderstanding, leading to ill-analyzed requirements. Furthermore, these requirements can be discussed in natural language, which heavily relies on legal jargon. Natural language leads to misinterpretation [24], and graphical notations can help communicate in interdisciplinary teams [31] as long as all stakeholders can understand the notation.

Following this line of thought, collecting evidence from practitioners and different regulations is critical for applying conceptual modeling techniques aiming to help stakeholders in this context. Firstly, ontologies can help the stakeholders achieve common ground when discussing DPRs. In addition, it also would identify the requirements across jurisdictions that are commonly conceptualized. For example, personal data may be conceptualized as data of an identified or identifiable living persona. Conceptual modeling would help with the communication process. Rather than discussing in natural language, the personal data flows and processing activities could be graphically represented to avoid misinterpretations when discussing and implementing specifications. The ontology would then define the relationship between classes, entities, and objects, making clear the ontological commitments of the conceptual modeling.

B. Research Design

To design and develop the OBI-PIA method, we have framed the project as both a design science approach per [32] and the conceptualization of the conceptual modeling method per [33], [34]. Next, we detail the design science stages of the ongoing research and then an overview of the main outcome of the process, the OBI-PIA method.

Re

1) Problem investigation: Identification of common regulatory data protection requirements worldwide: Regards the identification and analysis of common regulatory DPRs worldwide. Understanding the domain's current landscape and identifying stakeholders' needs and requirements is key for developing the artifact [32]. The objective of OBI-PIA is to be jurisdiction agnostic. Hence, we aim to identify common requirements and their legal conceptualization across countries. We aim to answer the knowledge questions about what are the common concepts and relationships for performing DPA across regulations? And what is the common procedure for performing DPA across regulations?. Eliciting these elements through interviews and analysis of regulatory documents will inform the design decisions of the treatment design stage. To answer this question, we follow a transdisciplinary approach: legal research methods for desk reviews, qualitative approaches (interviews) with lawyers, and specialized tools for analyzing regulations.

We started our research by conducting interviews with lawyers. Interviews are commonly used for eliciting and understanding requirements [35]. Our preliminary codebook for analyzing the interviews is based on analyses of peer-reviewed literature, jurisprudence, and sociological, legal studies to understand the spirit of the law [36], following a deductive approach [37]. This codebook was validated with five lawyers from different countries and years of expertise (ranging from PhD student, to experts with +10 years of experience). The codebook has been further refined according to the early analysis phase defined by [37], to better suit the data.

We are conducting semi-structured interviews with data protection legal experts to answer the knowledge question of our first phase [32]. The interview is designed to elicit the most common data protection regulatory principles, legal basis, data subjects' rights, actors and roles, duties, history of the regulation, data breach requirements, challenges in transborder flows, standardization, and communication processes. Additionally, we ask lawyers about their main concerns, requirements, and demands for their practice when working in interdisciplinary teams. This data will then be used for conceptualizing common DPRs as discussed in Section III-B. More information on the research method, anonymized data and details is available at https://tinyurl.com/espre2025 at the latest version.

For the interviews, due to amount of countries with data protection regulations, we focused on the G20, acknowledging that they may not fully represent all regions. Therefore, we randomly selected seven countries invited to the G-20 summit — the list consisted of all countries invited to G-20 summit since 2008 — plus Luxembourg (the host country). To include maximum variability for countries outside the EU, we aim to interview 4 to 6 lawyers (from academia, activism, regulator, and private practice) for saturation purposes [38] and 3 lawyers per EU country, which are part of the G20. In addition, we also included 4 lawyers from countries outside our sample that are part of the EU and African Union. Our final sample should be between 80 - 120 lawyers, depending on the saturation we achieve while interviewing.

The interviews are analyzed using a mixed deductive [37] and inductive approach, with thematic analysis and a coding book [37], [38] to explore the meanings and themes of data protection regulations and expert insights. This approach acknowledges the gap between written law and practice. We also conduct content analysis by quantifying specific elements for triangulation [32], [38]. To handle unforeseen data, we included the possibility of new codes (inductive research). Our coding unit are phrases.

Results from the interviews are complemented with comparing legal and policy literature about data protect. We seek to code phrases that could impose DPRs in their native language [37] and then perform a two-level synthesis to get to subcategories and categories of DPRs across regulations. In addition to common conceptualizations, we will also characterize the differences between regulations that could spark requirements change or conflicts regarding transborder data flows.

The outcome of this research process are the common conceptualizations across regulations based both on the legal documents and the practitioners' perspective.

At the moment of writing of this section, we have conducted 45 out of the planned 80-120 interviews. Our subjects have diverse backgrounds, with a range of 2 to 30+ years of professional experience across different continents, with gender parity. Their backgrounds range from data protection authority (or those working with them) to private practitioners, activists, academics, and PhD students. We achieved high data saturation levels — meaning no new substantial data has been found — in specific countries or areas at 4 interviews, as we are no longer discovering new data from these countries (Brazil, Chile, the African Union, China, and the EU). From a theoretical point of view, high levels (70%) of data saturation can be achieved at a small number (starting with 4 interviews [39], [40]), which is in line with what we have perceived. All data subjects have provided their granular consent to the interviews, and some have consented to make their interviews public at https://tinyurl.com/espre2025. Per default, we anonymize all data to lower our biases when analyzing the data.

2) Treatment design: Designing the OBI-PIA Method: In this stage, we address the problem of designing a conceptual modeling method for analyzing data protection requirements in an international context; we state our objective following Wieringa (2014) [32]:

- To improve the inclusion and analysis of regulatory transborder DPRs;
- by designing an evidence-based conceptual modeling method;

- that satisfies regulatory requirements from multiple jurisdictions;
- so that organizations can include data protection by design, help with documentation, and satisfy and demonstrate compliance with regulators.

The main artifact that needs to be designed is the conceptual modeling method.

To start the conceptualization of the constructs, we will use the data analyzed and gathered in the interviews and in the first phase, in a similar approach as [41] [32, pg.138]. Thus, the ontology is not the main artifact, but acts provides the semantic and syntact aspects of the modeling language [33].

3) Ontology development: This phase addresses the evidence-base part of the artifact. Using data from the problem investigation, we will explore how the conceptualization of the DPRs match. We expect that specific requirements may differ across jurisdictions and have different conceptualizations, while others may overlap.

Once we have the different conceptualizations, we will develop the domain ontology for those DPRs with similar conceptualizations. Previous research suggests that while some ontologies address some legal DPRs, international perspectives are lacking [21]. Likewise, there seems to be weak semantic grounding from the legal domain in existing RE ontologies, potentially compromising the conceptual fidelity [42]. From this starting point, developing the ontology for this domain — international regulatory DPRs — is a challenge. It must represent semantically the models of different stakeholders with different backgrounds from multiple jurisdictions.

Hence, we need to converge all these conceptual models into one ontology. To achieve this, we propose a three-fold policy: first, to thoroughly model a conceptualization of DPRs; second, to evaluate the conceptualization against the OECD principles (which are the cornerstone of most regulations [8]); and third, to discuss with lawyers their insight over the conceptualization. This approach will help reduce internal biases and include a legal perspective. However, the plan is not fail-proof, as there may be cases where there will be more than one model for a concept, both equally stringent. In those cases, we will discuss with lawyers what may be more semantically appropriate.

We aim to create a domain ontology applicable to various artifacts built upon the UFO foundational ontology. The ontology should enable common ground for regulatory DPRs [42]. The UFO paradigm provides the flexibility required for the domain, such as UFO-C, UFO-L, and UFO-A [42], [43]. The use of foundational ontology will help us systematically build a domain ontology. In our ontology, the foundational ontologies will be extended with concepts which are common across regulations, setting a common ground for a country-independent, first-level understanding of DPRs. These concepts then will be extended with country-specific concepts, materializing the different meanings and interpretations of such concepts in specific regulations. We will also try to strengthen the conceptual fidelity of the conceptualization and ontology by grounding our decision in empirical evidence. This decision also helps us make ontological commitment clear and transparent [23]. Additionally, it should represent the consensus of the domain community and can 'be supported by both theoretical and empirical evidence' [42]. This is why we carry out interviews in the problem investigation. Legal and software engineering experts will validate the ontology and abstraction in line with [32]. For validation, we will conduct focus groups or surveys with lawyers and requirements engineers [32], assessing users' effectiveness in information-retrieval tasks by asking them to define concepts from the ontology and rate their agreement with the proposed model.

4) Conceptual modeling: With the domain ontology developed, the artifact should develop the conceptual modeling method, including the modeling algorithms and procedures [33]. The method will be instantiated for DPIAs in transborder personal data flows contexts. Stakeholders from multiple backgrounds should be able to use OBI-PIA, which should maintain the richness and complexity of the data protection domain.

The understanding and usability of non-experts across cultures of the conceptual modeling method will be essential. There is research on measuring understandability of conceptual models [44], including regulatory DPRs [45]. OBI-PIA conceptual models should help with the documentation process of DPIAs. This documentation is crucial for satisfying compliance requirements and recording design decisions. Thus, we will review user-experience and technological acceptance model works regarding the method, working alongside stakeholders through focus groups and interviews. As this is the last phase, our plans are not fully developed yet — as we await the first two objectives — but this is what we envision.

To further develop certain aspect of OBI-PIA - such as the selection of specific graphical notations or views - we will follow the situational method of engineering of [46]. We plan to analyze existing models, concepts, and representations, identify possible re-usages, and triangulate how they work with our domain conceptualization. For example, existing artifacts from goal-oriented modeling, such as the information view from STS-ml [47], may be interesting to re-utilized. The concept of goals or agents (albeit adapted to a more 'legalese' language) may be valuable. The validation phase shall be done internationally, following the technical action research (TAR) approach. We plan to apply it with partnering organizations from at least two countries and regions to apply our proof-of-concept.

C. The proposed method: OBI-PIA

What we propose to help organizations with data protection compliance is a lightweight and usable conceptual modeling method based on an ontological conceptualization of the different national regulations. Such artifacts should help organizations develop their data protection impact assessments (DPIA) independently of the regulatory regime; *i.e.*, transcending any specificity to a particular legal method while preserving general and international legal validity.

The modeling language should help analyze and communicate regulatory DPRs for transborder data flows

within interdisciplinary teams. In other words, identify requirements that are common and can be reused in different jurisdictions, and detect those that may be particular to specific countries. To conceptualize the creation of the conceptual modeling method, we follow the paradigms presented in [34] and [33], to identify key aspects of the method.

1) Modeling language: would be an ontology-based modeling language to help communicate and analyze regulatory DPR from an international perspective. Following [33], [34], the modeling language comprises syntax, semantics, and notation. Starting with the semantics of the language, this item is provided by the ontology we are building. The ontology identifies which regulatory data protection requirements are common across regulations and their conceptualization. In particular, the ontology will define the constructs by building their concepts, relationships, roles, attributes, and semantic constraints, among others. The ontological commitments [42] of this semantics come from the analysis of the interviews, specialized grey literature, and legal academia.

The ontology will also help us work around the syntax of the modeling language, which should be well-specified and formulated, avoiding anti-patterns [42]. We will construct the artifact using the UFO paradigm and make explicit explanations of our meta-models.

For example, a data subject would be a person who has a role, as a person can also be a data protection officer or even act as a controller if they have a business. However, they cannot be a data subject and a data protection office/controller concurrently, as they have different rights and duties. In addition, a data subject should be a person currently living. Hence, the modeling language should be able to identify these three possible roles that a person can have, differentiate them, and communicate if they are living or deceased.

For the notation elements, we have conducted prior empirical research on what seems to work between lawyers and what seems to be difficult [6], [45]. For example, the social view of the STS-ml GDPR extension seems complicated for lawyers, or the concept of an actor over specific roles in data protection seems uninviting [45]. In this sense, we expect to re-use notation elements and diagrams that work with non-expert stakeholders.

2) Modeling procedures: We will build informal guidelines for this block by following the modeling method. DPIA guidelines provided by data protection authorities and professional organizations, along with best practices, will inspire these guidelines [48]. It will include vital steps, such as identifying which type of personal data is being handled, identifying retention periods and legal basis, and identifying the actors who handle data.

We also foresee constraints, which are representations of our semantic constraints. For example, an actor cannot simultaneously be a data subject and a controller. This constraint is because data subjects are commonly conceptualized as natural personas and controllers as legal personas (or a natural persona carrying out some business, which is a legal persona). Another constraint is that processed personal data must have a legal basis. Otherwise, the processing is not legal.

3) Modeling algorithms: A key aspect of the OBI-PIA methods is to identify unique local data protection requirements and those common across the world. As inputs, the stakeholders will create the models for personal data governance. The algorithm should first alert if vital attributes are missing or contradictory. For example, an actor is modeled as a data subject and controller. Or if the actors do not identify the geographical jurisdiction. Secondly, it should check if some information is missing, according to the best-practices standard, and suggest where to add more information. Once these models have been checked, the algorithm will help alert the stakeholders where there may be conflicting requirements between different regulations, which may deserve attention and seem in line.

For example, consent is a common legal basis for processing data, whereas a legitimate basis is not. Let us assume that organization A in Europe uses legitimate interests as its legal basis for processing data. Now, organization A wants to expand to Colombia, opening an office there and providing services. In Colombia, the only legal basis for data protection regulation is consent. Otherwise, other exceptional legal basis defined in other laws can be used (such as FinTech). The modeling algorithm should then alert the modeler about this, as they will have to change their legal basis for processing data in Colombia (from legitimate interest to consent/specialized law). If consent is chosen, there will be new requirements. This consent has to be free, informed, and unambiguously given, which are new requirements that organization A did not have in Europe, as the legal basis was different. Furthermore, the retention periods of such information also change, alongside the data subject's rights. As such, the modeling algorithm should be alert to these elements.

As a prototype and proof-of-concept, we plan to apply this modeling algorithm between two regulations. Our long-term vision is to scale up the OBI-PIA method to allow those interested in adding local regulations.

IV. Preliminary results

In the following section we present preliminary results of our problem investigation, with a first proposal on DPRs. These results are an early analysis of our deductive qualitative analysis research method discussed in Section III-B.

A. Preliminary analysis

As part of the deductive qualitative analysis, this article presents an early analysis of interviews [37], where we are immersing ourselves in the data and identifying high-level trends. In particular, a set of regulatory DPRs have started appearing systematically across various legal methods. These requirements seem to share similar conceptualization, albeit not equal. Other requirements and conceptualizations have shown critical differences.

a) Common conceptualization: A key common element is that most regulations have been influenced by the OECD Privacy Principles, serving as a cornerstone or inspiration for most regulations [8]. This was an expected finding, as such principles were created to facilitate transborder data flows [17].

Another requirement that has a high saturation of conceptualization is consent and personal data, as well as the response elements required for data breaches. Consent is commonly required to be free, informed, unambiguous, and specific in most regulations. Personal data is identified or identifiable data regarding a living natural persona. Only South Africa and India conceptualize it slightly differently, where the former includes organizations and the latter deceased persons. When prompted what lawyers would require/do in case of a data breach, every interviewee answered the same: understanding what personal data has been breached and their management policy, a clear incident response protocol, and identifying notification requirements for both authorities and data subjects. The interviewees stressed the importance of knowing the nature and data lifecycle policy of the personal data to assess the risk of the data breach.

Children's (or minors') personal data is also a transnational preoccupation. Most regulations seem to have some special provision for the data management of minors. However, these requirements vary. Therefore, identifying if an IS works with children is an important requirement across regulations.

According to our interviewees, other legal bases, apart from consent, seem to be shared among most modern regulations, such as vital interest or public interest. However, these bases may not exist in older 'unreformed' regulations. There is a similar situation regarding data subject rights, where some more modern rights, such as the right to erasure, have started to be included in modern regulations but not consistently.

All lawyers identified communication issues when discussing DPRs with non-lawyers (mainly engineers). They argued that lacking a shared understanding and common conceptualization of key elements led to misunderstanding and unfruitful discussions. Common examples included the definition of personal data, power asymmetries, vulnerability, notification requirements, among others. On the other hand, they would recognize that their lack of expertise in technical aspects hurts them in assessing if certain software specifications hurts or helps achieving DPRs. They stressed the necessity of a method that lawyers and engineers could use to discuss DPRs.

b) Differences: One vital difference between regulations has been the legal basis. Although some legal bases, such as the performance of a contract/legal obligations, public interest, and consent, seem to overlap between jurisdictions, others do not. Legitimate basis, for example, although popular in most modern regulations, does not appear to be present in all of them. Other jurisdictions have created legal bases that respond to their reality, which makes them unique. For example, jurisdictions like Colombia rely ordinarily on consent as a legal basis for personal data processing. On top there are domain-specific regulations that may provide other extraordinary legal bases for processing personal data, however consent should be the most used legal basis. How does, for example, this legal basis affect IS requirements? Consent needs to be specific, unambiguous, informed, and freely given. Therefore, organizations must have specific processes that show that they satisfy and document these conditions.

Regarding the roles of the actors involved in personal data processing and transfers, it is not clear-cut that the controller/processor is commonly conceptualized. Jurisdictions that have been heavily GDPR-influenced seem to have the same conceptualization. That is to say, controllers define the personal data management lifecycle, while processors carry out the activity on their behalf. However, other jurisdictions do not necessarily make differences between these roles (i.e., there is no processor role) or may create new roles (such as consent manager, like in India). Hence, understanding these differences will be key when doing the stakeholder mapping.

Similarly, fines and time-frames for reporting diverge between jurisdictions. Organizations must identify the time frames for reporting if they fall under the scope of a jurisdiction. In addition, reporting requirements are not the same everywhere. While some jurisdictions are required to report any data processing activity, others do not. Likewise, some regulations require reporting any data breach, whereas others will depend on the risk to the data subject. Hence, some interviewees have suggested that following the most constraining requirements in this aspect is good practice, as this will also help with compliance in more 'laxed' jurisdictions.

Regarding transborder personal data flows, there is no common conceptualization. Although several countries accept standard contractual clauses or consent, they do not necessarily pose the same requirements. For example, the GDPR indicates consent as an exceptional reason, whereas it would be the common legal basis in other regulations. We have noticed a trend of creating adequacy lists between newer regulations.

c) Lawyers' requirements: Most interviewees have recognized, to different degrees, that they struggle to talk with engineers or software developers. Most interviewees have highlighted that 'they speak different languages' when dealing with DPRs. Consequently, as they do not possess common ground on conceptualizing DPRs, what goals to achieve, and how to satisfy them, they have difficulties discussing DPRs.

On the reasons to explain this conflict, it seems there is no consensus on whether this difference in language is because of educational background or incentives. Some subjects have highlighted that this situation is a consequence of education, thus affecting their mental models on how DPRs are to be satisfied (in line with what [5] has suggested). Others have expressed that they believe software engineers have a good understanding of DPRs and the regulatory context. Therefore, their incentives for satisfying DPRs differ and impact their conceptualization; i.e., lawyers focus on minimizing risk for organizations and software engineers in building software.

Although it is not a new phenomenon that language plays a key role in interdisciplinary contexts, our interviewees have made it explicitly clear that they need an artifact to ground their language. One interviewee, in particular, told us they

had worked around conceptual modeling but struggled to understand how it worked.

B. Two-level game requirements

'It has a very specific logic, it has specific principles, which if someone sits down and reads them, they become quite understandable [...] The problem is that data protection law is not universal [...] But there is some logic, which is even if the specifications may be different between different jurisdictions, different countries, the logic does not change.' (Verbatim subject EU-N)

Understanding DPRs for personal data flows in international contexts requires a structured approach that identifies shared and divergent regulatory requirements. Inspired by Putnam's two-level game theory in international relations [49], we propose a preliminary conceptualization classification of DPRs into two levels. The first-level requirements are those commonly conceptualized across multiple jurisdictions, and the second-level requirements are specific to a regulation. This classification should help organizations analyze DPRs, to asses differences and commonalities across jurisdictions. Therefore, organizations can identify interoperable IS requirements with those that may need efforts to change the specifications to satisfy compliance with national regulations.

First-level requirements are DPRs that, while not identical, are conceptualized similarly across national regulations. For example, most data protection laws are inspired by the OECD Privacy Guidelines (as discussed in Section II), which is a first step for guiding the discussion of requirements. Similarly, personal data and special categories, consent, identification of the data subject as minor, and some subject's rights are also part of this level of requirements. By identifying these interoperable DPRs, organizations may not need to change the specification to achieve compliance across jurisdictions.

Second-level requirements, however, reflect DPRs that differ across regulations or may be unique to a specific jurisdiction. An example may be some data subjects' rights, such as the 'the right to be forgotten'. Our preliminary results show that this specific right is not necessarily present in most regulations and can differ in how it needs to be satisfied. For example, in Chilean regulation (both old and upcoming), there was an academic and policy debate about how should 'the right to be forgotten' be conceptualized, *à la européen* or in more in accordance to Chilean constitutional law [50]. Therefore, compliance with second-level requirements will depend on the context, identifying which exact DPRs are unique to that jurisdiction. Afterward, when these DPRs are identified, how to modify the specification to satisfy the requirements needs to be discussed.

This classification also raises the possibility of an intermediate level of requirements, which can include regional frameworks. Although this is still not part of our proposal, we envision that there may be an intermediate-level requirement. The OBI-PIA method will focus on the first level DPRs, which are those that are commonly conceptualized.

V. Threats, limitations and future work

Creating interdisciplinary artifacts can lead to excessive complexity, resulting in a high cognitive load for users. The primary aim of conceptual models is to provide abstraction, and merging disciplines may negatively impact this process. Particular attention should be paid to the balance of abstraction with semantic richness [42]. In the future, ontologies created by interdisciplinary teams could be compared to those developed within a single discipline. Additionally, the domain ontology formalization may be limited and not fully capture the domain's richness and vastness. Key details might be lost during abstraction, and not all mental models may be represented. While we will continually validate our work, these limitations must be acknowledged.

Given the project's global scope, there is a risk of missing some regulations. Focusing on the G-20 countries plus specific invitees may impact the external validity. We aim to increase the external validity by analyzing all the regulations worldwide through NLP em, which is under discussion. Future work could apply and analyze the method outside the scope of our selected countries.

Our primarily qualitative methods have inherent challenges, including case studies and interviews, which may have low external validity and introduce biases. To mitigate this, we will select interviewees from various countries and practice areas to tackle these issues, ensuring a sufficient sample size and saturation. We will also provide as much detail as possible when reporting our interviews while respecting privacy concerns and the consent options of the interviewee. Furthermore, there may be biases introduced by the coders. However, generalizations will be carefully approached, and results should be analyzed within the context. Interview questions were crafted based on theoretical analysis to ensure content validity and avoid socially-desirable answers. They were pre-tested between expert matter subjects and modified accordingly.

VI. Conclusion

This paper presents the OBI-PIA method, which helps stakeholders analyze regulatory data protection requirements in transborder data flows contexts. Capitalizing on conceptual modeling, the method should help different stakeholders discuss data protection, particularly in the contexts of DPIAS. By creating a domain ontology using the UFO paradigm in data protection, the conceptual modeling method should help stakeholders communicate, analyze, and discuss this type of requirements. Given the interdisciplinary and international nature of the domain, the project involves researchers from multiple disciplines and uses different research methods. The objective is to create a method that stakeholders from different jurisdictions and expertise can use to discuss and work around regulatory data protection requirements.

VII. Data availability statement

Codebooks, interview questions and transcripts consented to being public, are available at https://tinyurl.com/espre2025

or https://zenodo.org/records/15676602. We share all data compatible with interviewees' privacy.

REFERENCES

[1] R. Gellert and S. Gutwirth, "The legal construction of privacy and data protection," *Computer Law & Security Review*, vol. 29, no. 5, pp. 522–530, 2013.

[2] T. Breaux and T. Norton, "Legal accountability as software quality: A u.s. data processing perspective," in *International Requirements Engineering Conference (RE)*. IEEE, 2022.

[3] T. D. Breaux and A. I. Antón, "A systematic method for acquiring regulatory requirements: A frame-based approach," *RHAS-6, Delhi, India*, 2007.

[4] T. Breaux and A. Antón, "Analyzing regulatory rules for privacy and security requirements," *IEEE transactions on software engineering*, vol. 34, no. 1, pp. 5–20, 2008.

[5] I. Hadar, T. Hasson, O. Ayalon, E. Toch, M. Birnhack, S. Sherman, and A. Balissa, "Privacy by designers: software developers' privacy mindset," *Empirical Software Engineering*, vol. 23, no. 1, pp. 259–289, 2018.

[6] C. Negri-Ribalta, R. Noel, O. Pastor, and C. Salinesi, "An empirical study on socio-technical modeling for interdisciplinary privacy requirements," in *CooppIS*. Springer, 2023.

[7] F. Casalini, J. Lopez-Gonzalez, and T. Nemoto, "Mapping commonalities in regulatory approaches to cross-border data transfers," *OECD Trade Policy Papers*, no. 248, 05 2021.

[8] O. for Economic Co-Operation and Development, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," Tech. Rep., 1980.

[9] O. Babalola, "Transborder flow of personal data (tdf) in africa: Stocktaking the ills and gains of a divergently regulated business mechanism," *Computer Law & Security Review*, vol. 52, p. 105940, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0267364924000074

[10] O. for Economic Co-operation and Development, "Report on the implementation of the oecd privacy guideline," 2023.

[11] ——, "Cross-border data flows: Taking stock of key policies and initiative," 2022.

[12] D. J. Solove, "A taxonomy of privacy," *University of Pennsylvania law review*, 2006.

[13] G. G. Fuster, *The emergence of personal data protection as a fundamental right of the EU*. Springer Science & Business, 2014, vol. 16.

[14] P. Jurcys, M. C. Compagnucci, and M. Fenwick, "The future of international data transfers: managing legal risk with a 'user-held'data model," *Computer Law & Security Review*, vol. 46, p. 105691, 2022.

[15] L. A. Bygrave, "The 'strasbourg effect'on data protection in light of the 'brussels effect': Logic, mechanics and prospects," *Computer law & security review*, vol. 40, p. 105460, 2021.

[16] A. Mattoo and J. P. Meltzer, "International data flows and privacy: The conflict and its resolution," *Journal of International Economic Law*, vol. 21, no. 4, pp. 769–789, 12 2018. [Online]. Available: https://doi.org/10.1093/jiel/jgy044

[17] O. for Economic Co-operation and Development, "Oecd declaration on transborder data flows," 1985.

[18] ——, "Oecd guidelines on the protection of privacy and transborder flows of personal data," 2002.

[19] E. J. Novotny, "Transborder data flow regulation: Technical issue of legal concern," *Computer/LJ*, vol. 3, p. 105, 1981.

[20] G. B. Herwanto, F. J. Ekaputra, G. Quirchmayr, and A. M. Tjoa, "Towards a holistic privacy requirements engineering process: Insights from a systematic literature review," *IEEE Access*, 2024.

[21] C. Negri-Ribalta, M. Lombard-Platet, and C. Salinesi, "Understanding the gdpr from a requirements engineering perspective—a systematic mapping study on regulatory data protection requirements," *Requirements Engineering*, pp. 1–27, 2024.

[22] M. Gharib, P. Giorgini, and J. Mylopoulos, "Ontologies for privacy requirements engineering: A systematic literature review," *arXiv preprint arXiv:1611.10097*, 2016.

[23] G. Guizzardi and N. Guarino, "Explanation, semantics, and ontology," *Data & Knowledge Engineering*, p. 102325, 2024.

[24] M. Glinz and S. A. Fricker, "On shared understanding in software engineering: an essay," *Computer Science-Research and Development*, vol. 30, pp. 363–376, 2015.

[25] C. Kalloniatis, E. Kavakli, and S. Gritzalis, "Addressing privacy requirements in system design: the pris method," *Requirements Engineering*, vol. 13, no. 3, pp. 241–255, 2008.

[26] B. Esteves and V. Rodríguez-Doncel, "Analysis of ontologies and policy languages to represent information flows in gdpr," *Semantic Web*, vol. 15, no. 3, pp. 709–743, 2024.

[27] M. Robol, M. Salnitri, and P. Giorgini, "Toward gdpr-compliant socio-technical systems: modeling language and reasoning framework," in *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer, 2017, pp. 236–250.

[28] S. Agostinelli, F. M. Maggi, A. Marrella, and F. Sapio, "Achieving gdpr compliance of bpmn process models," in *Information Systems Engineering in Responsible Information Systems: CAiSE Forum 2019, Rome, Italy, Proceedings 31*. Springer, 2019.

[29] D. Torre, M. Alferez, G. Soltana, M. Sabetzadeh, and L. Briand, "Modeling data protection and privacy: application and experience with gdpr," *Software and Systems Modeling*, vol. 20, pp. 2071–2087, 2021.

[30] S. Sheth, G. Kaiser, and W. Maalej, "Us and them: A study of privacy requirements across north america, asia, and europe," in *Proceedings of the 36th International Conference on Software Engineering*, ser. ICSE 2014. New York, NY, USA: Association for Computing Machinery, 2014, p. 859–870.

[31] D. Damian and J. Chisan, "An empirical study of the complex relationships between requirements engineering processes and other processes that lead to payoffs in productivity, quality, and risk management," *IEEE Trans. Software Eng.*, vol. 32, pp. 433–453, 07 2006.

[32] R. J. Wieringa, *Design science methodology for information systems and software engineering*. Berlin, Heidelberg: Springer, 2014.

[33] N. Visic, H.-G. Fill, R. A. Buchmann, and D. Karagiannis, "A domain-specific language for modeling method definition: From requirements to grammar," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2015, pp. 286–297.

[34] H.-G. Fill and D. Karagiannis, "On the conceptualisation of modelling methods using the adoxx meta modelling platform," *Enterprise Modelling and Information Systems Architectures*, vol. 8, no. 1, pp. 4–25, 2013.

[35] K. Pohl and C. Rupp, *Requirements Engineering Fundamentals: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB Compliant*, ser. Rocky Nook computing. Rocky Nook, 2015.

[36] P. Chynoweth *et al.*, "Legal research," *Advanced research methods in the built environment*, vol. 1, 2008.

[37] S. T. Fife and J. D. Gossner, "Deductive qualitative analysis: Evaluating, expanding, and refining theory," *International Journal of Qualitative Methods*, vol. 23, p. 16094069241244856, 2024.

[38] E. R. Babbie, *The practice of social research*. Cengage learning, 2020.

[39] M. Hennink and B. N. Kaiser, "Sample sizes for saturation in qualitative research: A systematic review of empirical tests," *Social science & medicine*, vol. 292, p. 114523, 2022.

[40] G. Guest, E. Namey, and M. Chen, "A simple method to assess and report thematic saturation in qualitative research," *PloS one*, vol. 15, no. 5, p. e0232076, 2020.

[41] Z. Racheva, M. Daneva, K. Sikkel, A. Herrmann, and R. Wieringa, "Do we know enough about requirements prioritization in agile projects: Insights from a case study," in *2010 18th IEEE International Requirements Engineering Conference*, 2010, pp. 147–156.

[42] G. Guizzardi, "Ontological foundations for structural conceptual models," 2005.

[43] C. Griffo, J. P. A. Almeida, and G. Guizzardi, "Conceptual modeling of legal relations," in *Conceptual Modeling: 37th International Conference, ER 2018*. Springer, 2018.

[44] C. Houy, P. Fettke, and P. Loos, "Understanding understandability of conceptual models–what are we actually talking about?" in *Conceptual Modeling: 31st International Conference ER 2012, Florence, Italy, October 15-18, 2012. Proceedings 31*. Springer, 2012.

[45] R. Velasquez, C. Negri-Ribalta, R. Noel, and O. Pastor, "Exploring understandability in socio-technical models for data protection analysis: Results from a focus group," in *International Conference on Conceptual Modeling*. Springer, 2023, pp. 263–273.

[46] B. Henderson-Sellers, J. Ralyté, P. J. Ågerfalk, and M. Rossi, "Situational method engineering," 2014.

[47] F. Dalpiaz, E. Paja, and P. Giorgini, *Security requirements engineering: designing secure socio-technical systems*, Cambridge, Massachusetts, 2016.

[48] A. . D. P. W. PARTY, “Guidelines on data protection impact assessment (dpia) and determining whether processing is “likely to result in a high risk” for the purposes of regulation 2016/679,” 2017. [Online]. Available: https://ec.europa.eu/newsroom/article29/items/611236

[49] R. D. Putnam, “Diplomacy and domestic politics: the logic of two-level games,” in *International organization*. Routledge, 2017, pp. 437–470.

[50] L. Ortiz Mesías and P. Viollier, “Repensando el derecho al olvido y la necesidad de su consagración legal en chile,” *Revista Chilena de Derecho y Tecnología*, vol. 10, no. 1, p. pp. 77–109, jun. 2021.