MDPI

Article

DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue based on GDPR

Paul Ryan [1,2,4]*, Rob Brennan [2,4] and Harshvardhan J. Pandit [3,4]*

[1] Uniphar PLC, Dublin, Ireland
[2] Dublin City University, Dublin, Ireland
[3] Trinity College Dublin, Dublin, Ireland
[4] ADAPT SFI Research Centre, Ireland
* Correspondence: P. Ryan paul.ryan76@mail.dcu.ie; H. J. Pandit pandith@tcd.ie

Abstract: GDPR requires Data Controllers and Data Protection Officers (DPO) to maintain a Register of Processing Activities (ROPA) as part of overseeing the organisation's compliance processes. The ROPA must include information from heterogeneous sources such as (internal) departments with varying IT systems and (external) Data Processors. Current practices use spreadsheets or proprietary systems that lack machine-readability and interoperability, presenting barriers to automation. We propose the Data Processing Catalogue (DPCat) - a specification based on DCAT-AP and the Data Privacy Vocabulary (DPV) for the representation, collection and transfer of ROPA information, as DCAT catalogues in a machine-readable and interoperable manner. DPCat represents a comprehensive semantic model developed from GDPR's Article 30 and an analysis of the 17 ROPA templates from EU Data Protection Authorities (DPA). To demonstrate the practicality and feasibility of DPCat, we represent the European Data Protection Supervisor's (EDPS) ROPA documents using DPCat, verify it with SHACL to ensure the correctness of information based on legal and contextual requirements, and produce reports and ROPA documents based on DPA templates using SPARQL. DPCat supports a data governance process for data processing compliance to harmonise inputs from heterogeneous sources to produce dynamic documentation that can accommodate differences in regulatory approaches across DPAs and ease investigative burdens toward efficient enforcement.

Dataset: https://w3id.org/dpcat/repo

Dataset License: CC-BY

Keywords: GDPR; data governance; semantic-web

Citation: Ryan, P; Brennan, R; Pandit H. J. DPCat. *Information* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Many organisations are complex entities that perform heterogeneous processing on diverse personal data, often organised using multiple organisational units or outsourced processing partners and sometimes under the jurisdiction of multiple data protection authorities (DPAs). Under the EU's General Data Protection Regulation (GDPR), organisations that act as a 'Data Controller' are obliged to create and maintain a "Register of Processing Activities (ROPA)" as a comprehensive record of personal data processing activities carried out under their responsibility (GDPR Art. 30). The ROPA, as described in GDPR, is a temporal snapshot of the organisation's practices and is the point of initiating communication or investigation regarding compliance, such as with a DPA. It is thus an important part of the organisation's processes related to ensuring and documenting its compliance.

In practice, organisations struggle to keep accurate and up to date ROPAs [1]. They often fail to integrate the maintenance and management of the Register of Processing Activities into their day to day operations [1]. This can result in a breakdown in the

GDPR Accountability Principle[1] as there is a lack of clarity as to the who, how, and when of updating the ROPA. To assist organisations with their ROPA-related duties, DPAs have provided guidance and templates that intend to ease the task of understanding requirements and harmonise the documentation through commonly-used formats and environments - such as spreadsheets [2,3]. In providing these templates, DPAs indicate what can be considered 'good practice' regarding what information should be documented within a ROPA. However, despite being based on a common legal obligation (GDPR Art.30), there is variance in the templates provided by DPAs where additional fields (not in the GDPR) are also encouraged to be documented [2]. An organisation operating in multiple jurisdictions is thus tasked with consolidating differing requirements from each DPA as either a distinct set of ROPA documents or a single combined one.

Further, the exercise of gathering the information necessary to create a ROPA is not a one-off activity [4] as there may be several data sources both internally (e.g. departments) [5] and externally (e.g. Data Processors) [5,6]. Therefore, ROPA creation requires communication between these distinct units to collate information pooled from 'heterogeneous sources' into a singular location to produce a ROPA. This necessitates some form of information management process for the tasks associated with documents, such as - reading or viewing, writing all or parts of it, exchanging them between relevant stakeholders, and ensuring their correctness and availability (e.g. backups or version control).

To address such requirements, the market vendors offer dedicated solutions for ROPA management, often as part of a larger suite of GDPR compliance tools [7].This follows the increasing trend of organisations adopting Regulatory Technology (RegTech) [8] to assist with legal compliance and requirements. The utilisation of ROPA is poised to be an important and key feature given its importance in the GDPR compliance processes.

However, these RegTech solutions are primarily centralised, proprietary, and emphasise custom processes that cannot be utilised outside of vendor-defined use-cases. In particular, the information being exchanged between internal and external stakeholders has been poorly researched in academia and commercial offerings (see section 2) despite the need for shared business and regulatory taxonomies for facilitating semantic interoperability [9] between stakeholders to identify feasible and compliant software solutions for data protection and privacy regulations [10,11].

There is a lack of ROPA related explorations in academic research, with existing efforts limited to early-stage work involving Enterprise Architecture models [12] or data [13]. For larger projects that have focused on GDPR compliance with explicit requirements regarding non-proprietary technologies and focusing on interoperability (e.g. semantic web), there is a distinct absence of research addressing ROPA-related tasks despite overlapping with the same information requirements. In terms of ongoing work, the ONTOROPA project [11] proposes building a semantics-based ROPA with blockchain-based trust guarantees.

We propose an approach to solving these challenges whereby we identify what data is required to complete ROPA, who are the ROPA stakeholders, how do they utilise ROPA and what are the required information flows requiring interoperability and machine-readability of ROPA. To address the identified challenges and their solutions, we present our work based on the following research objectives:

RO1 Identify information and information flows relevant for a ROPA in terms of stakeholders based on GDPR and EU DPAs guidelines and templates

RO2 Develop a machine-readable specification for representing and exchanging ROPA relevant information in an interoperable manner

RO3 Specify a mechanism for using developed machine-readable formats for aggregation, querying, validation, and exporting of information based on identified ROPA-related information flows.

[1] GDPR Article 5.2

Our previous work on this topic consisted of creating a semantic model of ROPA [5]. In this, we evaluated the GDPR and 6 DPAs templates and guidelines to identify a set of concepts required for the representation of ROPA related information and proposed its formulation as a '*common semantic model*' for representing commonality across the EU. We utilised the Data Privacy Vocabulary (DPV) [14], developed by the W3C Data Privacy Vocabularies and Controls Community Group[2] (DPVCG), as a vocabulary for representing identified concepts. We found and reported missing concepts to DPVCG, which subsequently extended the DPV with our contribution. We further developed our common semantic model into a proposal for establishing a 'Data Processing Catalogue (DPCat) [15] that utilises DCAT Application profile for data portals in Europe (DCAT-AP) [16] specification, itself based on the DCAT v2 standard, to represent the ROPA related information in the form of 'datasets' and 'catalogues' that could be maintained, used, and shared consistently.

This article expands on our prior work to provide a more complete and feasible solution for establishing a common machine-readable and interoperable mechanism for a common representation of ROPA. We extended the common semantic model to incorporate ROPA templates from all EU DPAs (17 of 31 DPAs have published templates) and updated the DPCat specification and the DPV to support representing this information. To demonstrate its practical application and usefulness, we applied the DPCat specification to ROPA documents published by the European Data Protection Supervisor (EDPS) for each identified use case (see Section 6). Finally, we go beyond state of the art by demonstrating the potential of our solution in realising the EU's 'Data Spaces' vision [17] by creating 'compliance-related specifications' that support representation (RDF), querying (SPARQL), validation (SHACL), and exchange (DCAT+DPV) of information.

The principal contributions of this paper are summarised as follows:

1. Use-cases exploring ROPA data governance and stakeholders (*RO1*)
2. A Common Semantic Model for ROPA (CSM-ROPA) representing information requirements from EU DPAs (*RO2*)
3. Data Processing Catalogue (DPCat) specification for representing and exchanging ROPA related information and provenance (*RO2*)
4. Demonstration of representation, querying, validation, and exchange of ROPA related information using DPCat and semantic web technologies (*RO3*)
5. Discussion on the practicality and application of DPCat as a 'common mechanism' for exchanging compliance information

The remainder of the paper is structured as follows: Section 2 discusses the state of the art and related work, and section 3 describes the development of the Common Semantic Model for ROPAs (CSM-ROPA) development. In Section 4, we discuss ROPA Information flows and Data Governance requirements for ROPA. Section 5 describes the DPCat data processing catalogue to enable ROPA information sharing, aggregation and querying for ROPA stakeholder interoperability. Section 6 provides an application use case to demonstrate the practicality and feasibility of DPCat. The remainder of the paper discusses the impact of our approach on real-world use cases based on enabling better automation and tooling for regulatory compliance and critically for authorities to ease investigative burdens towards effective enforcement, and we provide our conclusions and recommendations for future work.

2. State of the Art and Related Work

This section presents an overview of relevant work specifically regarding modelling, creation, and maintenance of ROPAs, and tangentially regarding GDPR related machine-readable and interoperable information management and compliance processes.

[2] H. J. Pandit chairs DPVCG and is the editor of DPV https://www.w3.org/community/dpvcg/

2.1. Information management solutions for ROPA

The International Association of Privacy Professionals (IAPP), the largest global community for privacy and data protection professionals, reported that 65% of organisations relied on spreadsheets or completely manual solutions to maintain their ROPAs [18]. Another IAPP report found 169 vendors supplying ROPA related information management services and software in 2020 [19]. This practice can be seen as being reflective of the prevalence of maintaining compliance-related information in 'manual tools', such as spreadsheets, without using technological solutions that operate on them. Instead, the exception to these is proprietary solutions offered by vendors, such as One Trust, Data Grail and Transcend. These privacy vendors have seen the importance in offering tools that enable integration with their solutions , however such integrations link into their locked ecosystem without the ability for organisations to control their data or to move it to an alternate technology provider. DPAs, in reaction to existing common practices, also provide spreadsheet templates that encourage use of manual or vendor-specific solutions.

One of the main failings of organisations regarding ROPAs is devolving its maintenance to their DPO and not having active involvement in its upkeep [1]. Best practice for ROPA suggests complete involvement of stakeholders in the ongoing maintenance of ROPAs [4] to provide the DPO with an accurate and up to date view of personal data processing carried out by that organisation [20]. This means that we need processes that assist DPOs and enable the engagement of stakeholders in the upkeep and review of ROPAs. From this, we conclude there is demand for automation through technological solutions and that the market is responding to such needs with commercial offerings.

Further, the Future of Privacy Forum (FPF) [21] reports that privacy and data protection technology providers also face a significant obstacle regarding lack of common terminology [22]. Therefore, the demand for automation and technologies should be accompanied by requirements for common mechanisms and terminologies that can operate across processes and stakeholders and establish standardised mechanisms within the ecosystem. We can look towards the manufacturing and finance sectors, where the harmonisation achieved through commonality and standardisation has improved regulation and value chains [9].

In response, we identify information and information flows relevant for ROPA governance based on GDPR and DPA guidelines and templates (*RO1*), and provide a common terminology for representing ROPA to overcome the lack of a common terminology (*RO2*).

2.2. GDPR compliance approaches using machine-readable metadata

In contrast to market reaction, ROPA as a topic has received little attention within the academic and research communities despite evidence of a broad category of approaches addressing GDPR compliance. Rozenthal et al. [23] propose 'Enterprise Architecture' as an ideal source for representing processing activities and technologies in an organisation. This is supported by Burmeister et al., who also investigate how Enterprise architecture can provide a DPO with insights on organisational data processing activities concerning GDPR compliance [24]. Enterprise Architecture has further explored sources of ROPA related information in the next subsection.

The ONTOROPA project [11] proposes using semantic web ontologies and knowledge graphs for representing ROPA related information, and using a blockchain to certify its integrity and authenticity. To address such challenges, research efforts at producing common terminology using semantic web vocabularies and ontologies have been developed [14,25]. Other approaches utilise such vocabularies to construct 'legal knowledge bases' and utilise them for compliance evaluation and monitoring, which can help harmonise and facilitate a joint approach between legal departments and other stakeholders to identify feasible and compliant solutions around data protection and privacy regulations [10].

Several Semantic-based projects provide Ontologies, vocabularies, and policy languages that can be utilised to represent GDPR concepts. These mainly focus on terms referenced in GDPR rights and obligations. Most projects focus on legal compliance evaluation rather than deployment and interoperability. They do not consider the critical aspect

of how the information required is maintained or generated within/by organisations, and the stakeholders and information flows involved in this process. Some notable outputs for this are: BPR4GDPR's IMO [26], GDPRov [27], GConsent [28], DPV [14], GDPRtEXT [25], SPECIAL's ontologies [29] and PrOnto [30].

BPR4GDPR (Business Process Re-engineering and functional toolkit for GDPR compliance) [26,31] is a relevant ontology-based compliance methodology used to dictate and evaluate processes. It is based on advanced process mining from event logs of IT systems to discover, monitor, and improve processes without pre-modelling the processes before mining. BPR4GDPR thus creates a novel process monitoring architecture with constraints for conformance checking and automated evolution of processes to satisfy the rules. It will take significant trials and development before the widespread deployment of advanced techniques like this in conventional organisations.

GDPRov [27], GConsent [28], and SPECIAL [29] provide ontologies for expressing GDPR related concepts, but not incorporate ROPA requirements. GDPRtEXT [25] provides a vocabulary of GDPR concepts, of which some concepts relate to ROPA (GDPR Art.30). PrOnto [30] provides concepts regarding data types, documents, agents and roles, purposes, legal bases (and more) - but is not available for reuse. DPV [14] also provides concepts regarding data categories, purposes, legal bases (and more), represents a community consensus, and is available for reuse.

These existing efforts, specifically DCAT(-AP) and DPV, provide the basis on which we develop a machine-readable specification to represent and facilitate the exchange of ROPA relevant information in an interoperable manner (*RO2*), and to utilise it for aggregation, querying, validation, and exporting of information based on identified ROPA-related information flows using the semantic web technologies (*RO3*).

3. A Common Semantic Model for ROPAs (CSM-ROPA)

Despite a ROPA being based only on requirements established by GDPR Art. 30, our prior work found variance amongst ROPAs templates provided by 6 DPAs in terms of what information needed to be documented. The additional fields were related to what the DPAs considered best practices to assist organisations in collecting and representing information from their various business processes. We harmonised the requirements from different templates to construct a 'common semantic model' for ROPA (CSM-ROPA) to enable the representation of all DPA-specified ROPA information [2]. We then represented these information requirements through concepts from the Data Privacy Vocabulary (DPV) [14] to provide an interoperable machine-readable vocabulary that can act as a mediation mechanism between stakeholders and tools operating on ROPA and associated compliance processes. In this section, we present results from our extended work where we analysed and incorporated ROPA templates from all EU DPAs to create a single (and truly) 'common semantic model' for ROPA and represented it using DPV to provide a consistent and interoperable specification for representing ROPA and its relevant information.

3.1. Analysis of DPA ROPA Templates

The GDPR has 31 DPAs[3] representing nations and member states from the EU and the EFTA EEA. Each DPA provides guidance regarding ROPA based on its basis in GDPR Art.30, and some DPAs also provide templates to assist organisations with maintaining their ROPA documents. In our prior work analysing 6 DPA templates [2], we found that the DPA ROPA templates go beyond the GDPR Art.30 requirements, are not consistent (with other DPA templates), and represent a challenge in producing a *collective understanding* of what information is required for maintaining a ROPA.

[3] Based on EDPB membership, 31 DPAs from 27 EU Member States, the EDPS, and 3 additional members comprising the EFTA EEA states. The German regional DPAs were considered part of the central DPA

In this work, we expanded the analysis to all 31 DPAs, and found 17 DPAs provided ROPA templates varying[4] in language and content. On these, we performed term extraction, semantic analysis, term frequency enumeration, de-duplication, and antonym/homonym identification. Templates with minimal information restricted their contents for conforming with GDPR Art.30. Some templates, such as those provided by Belgian and Greek DPAs, were extensive in fields beyond what the GDPR or other DPAs suggested.

The exercise, carried out over 2020-2022, yielded 47 unique concepts representing information to be recorded in a ROPA. Of these, 18 concepts were related to the requirements defined in GDPR Art.30, and the rest (29 concepts) were either supplementary to these or added by DPAs[5]. An overview of the exercise is presented in Appendix A, which shows the identified concepts and their relevance to each DPA template analysed.

3.2. Developing a Semantic Model for ROPA using DPV

In our previous work [2,3], we utilised DPV to represent terms identified from ROPA templates as machine-readable and interoperable concepts for use in information management and compliance-based approaches. Through this, we proposed a 'Common Semantic Model for ROPA' (CSM-ROPA). In this section, we describe our work in expanding the CSM-ROPA to cover additional requirements and concepts identified from the analysis of DPA ROPA templates and incorporate the updates made to DPV.

The DPV provides a semantic vocabulary consisting of hierarchical taxonomies of concepts relevant to GDPR, such as personal data, purposes, processing operations, technical and organisational measures, legal bases, and entities. We chose DPV as it provides the most comprehensive vocabulary for our purposes, is open and accessible, has ongoing development and mechanisms to submit contributions, and is familiar to the authors.

The process of representing identified concepts using DPV used the methodology [32]: for each term, we identified whether the DPV contained the (semantically) exact concept - which we call an '*exact match*', failing which we looked for the closest relevant term(s) which could be used as a substitute - called a '*partial match*', and if any existing term could not represent the term - we considered it a '*new term*' to be proposed to the DPVCG for inclusion in the DPV. Of the 47 unique concepts found through ROPA templates analysis, we found 44 exact matches, one partial match, and two new terms proposed and added to the DPV. Appendices A and B provide an overview of this outcome.

The output of this was the CSM-ROPA consisting of 47 concepts covering information requirements from GDPR and DPA templates for representing a ROPA. CSM-ROPA, through the use of DPV concepts, provides the ability to express a ROPA as a machine-readable and interoperable 'graph' that can be utilised in technological solutions for automating processes associated with ROPA and GDPR compliance. The CSM-ROPA data and analysis are available online[6].

4. Information and Data Governance for ROPA

The CSM-ROPA, described in the previous section, enables the representation of a ROPA in a machine-readable and interoperable manner and covers information requirements from the GDPR and DPA ROPA templates. However, a ROPA is not a single document in practice but is a related set of evolving information that must be periodically collected and maintained. The information required for maintaining a ROPA thus may have one or more internal sources, such as a department, unit, or assigned person - where such 'organisational units' provide data about their respective processes and activities. A ROPA may also have one or more external sources - such as processors, contractors, vendors - where such 'external entities' provide the information required for establishing records of agreed activities and assurance of compliance obligations.

[4] Of 17 DPA templates, 5 used English. We used Google Translate to convert the rest to English and manually ensured consistency in translation between templates regarding terms used.

[5] We could not discern source or basis in law (EU or national) for concepts added by DPAs

[6] https://w3id.org/dpcat/csm-ropa

The ROPA provides the DPO with an important overview of the organisation's practices [20], and is part of the DPO's obligations regarding compliance (GDPR Art. 39) [33]. This requires communication between internal stakeholders such as units or departments, and external stakeholders such as DPAs, auditors, and certification bodies - to collate necessary information for ROPA governance.

We present five use cases[7] that explore the key stakeholders and their roles regarding the '*heterogeneous sources*' in ROPA related data governance. This follows the methodology from prior work [6] regarding identifying stakeholders and information flows related to GDPR compliance and establishing the utility of developing machine-readability and semantic interoperability mechanisms based on it.

In our analysis, we considered the DPO as the nominated entity with responsibility within an organisation to oversee the ROPA related processes as per the obligations from GDPR (Art. 39). From this perspective, we explore possible combinations based on the existence or involvement of specific stakeholders and their effect on the DPO's duties to collect and maintain ROPA related information. We also considered a Data Controller as the primary type of organisation despite a Data Processor being required to maintain ROPA and involve a DPO as a stakeholder. The Data Controller's use-cases are more complex than a Data Processor's, and a solution satisfying a controller's ROPA requirements can be trivially modified for use by a processor.

This exercise concludes with an argument for the expression of ROPA related information in a machine-readable and interoperable format. Section 6 then presents DPCat as our solution to communicate or exchange ROPA relevant information between stakeholders and can assist in the automation of compliance processes.

4.1. Use Case U1: Data Controller

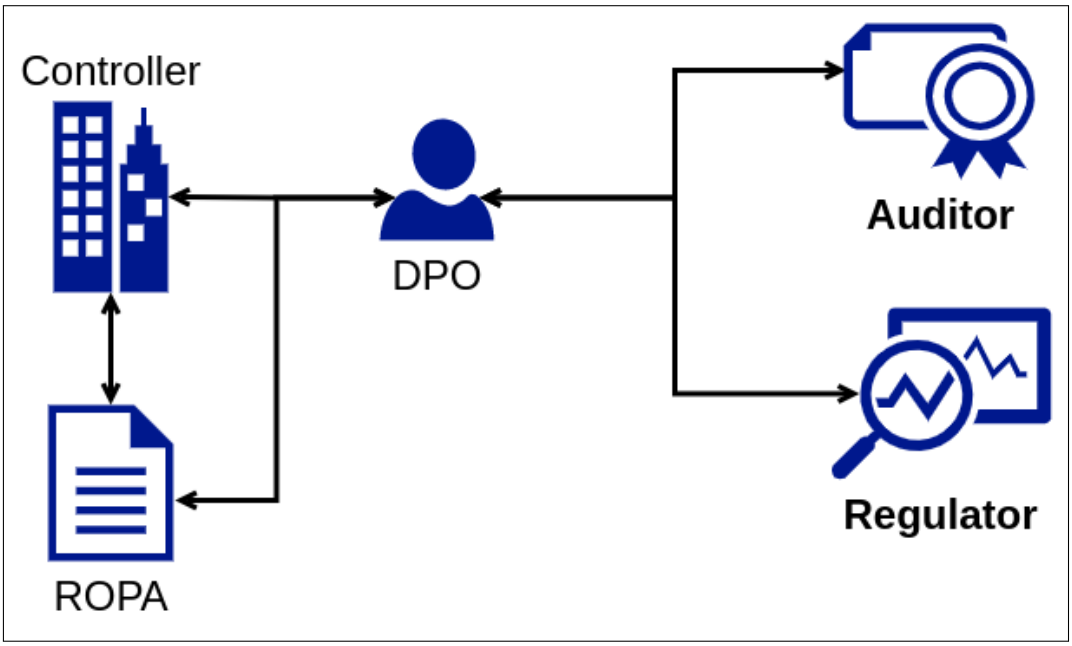

Figure 1. Basic generation of legal requirement ROPA

This use case, illustrated in Figure.1 represents a single Data Controller that maintains a ROPA (GDPR Art.30), for which it identifies and documents relevant processing activities conducted under its responsibility. In addition, as best practice, the controller must assess guidelines and templates provided by relevant DPA(s) and adapt its documentation processes accordingly to meet any additional suggestions or requirements. A ROPA produced

[7] In this, we relied on P. Ryan's experience as an active DPO for over 30 legal entities

by a Controller is utilised by its DPO as part of the responsibility to oversee compliance. The ROPA may also be accessed by a DPA or an auditor (e.g. a certification body) as part of their correspondence with the controller or an investigation or auditing process.

The information flows between these stakeholders can involve: (i) A ROPA that conforms to GDPR Art.30 requirements; (ii) A ROPA that conforms to a DPAs guidelines and templates; (iii) Provenance, e.g. ROPA issuer, timestamps, contact details; and (iv) A selective part of ROPA, e.g. temporal period, specific processing activities.

4.2. Use Case U2 Data Controller with Internal Organisational Units

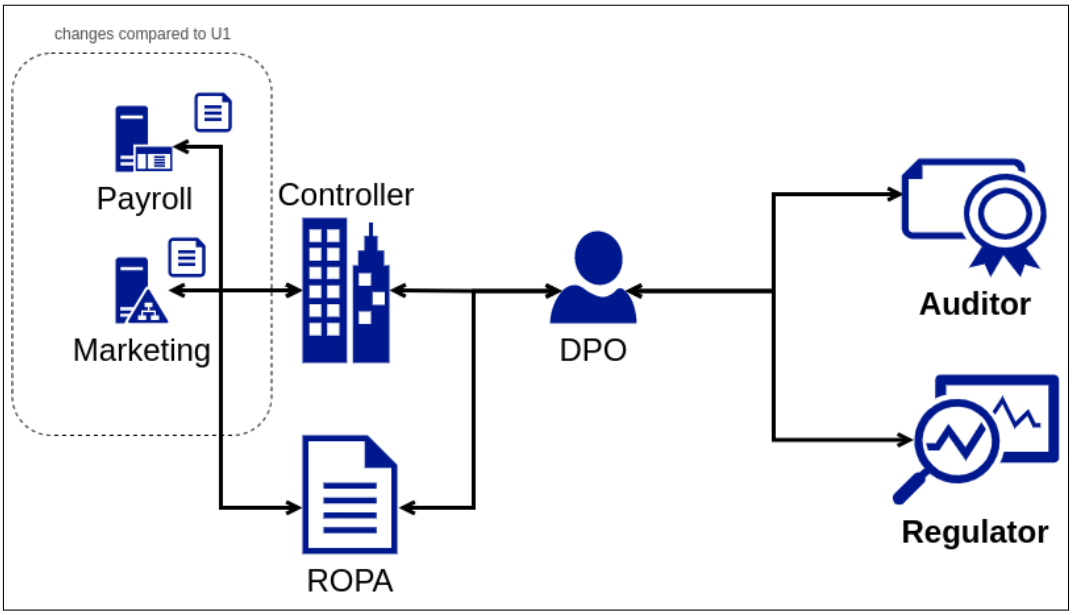

Figure 2. Organisational Units updating and maintaining ROPA

The second use-case, presented in Figure.2, expands U1 with internal information flows through four *'organisational units'* or departments: Marketing, Human Resources, IT Services and Web Services, where relevant data for generating a ROPA must be collated from them into a common location [4]. U2 also involves potential follow-ups with each unit regarding maintenance of records per-department, and establishing 'points of contact' and 'responsible entity'. The external information flows, i.e. DPAs and auditors, stay the same as internal units are not separate legal entities subject to direct investigation.

The key information flows between these stakeholders, in addition to those in U1, may involve: (i) Complete or partial ROPA information for each internal organisational unit; (ii) Provenance, e.g. department as issuer, point of contact or responsible entity, timestamps, contact details; (iii) Collation of department information into a common ROPA for external stakeholders; and (iv) A selective part of ROPA, e.g. specific department.

4.3. Use Case U3: Data Controller with Data Processors

The third use case, illustrated in Figure.3, has additional information flows where the controller and its DPO collect relevant information from appointed processors. In cases where a processor is common to all departments or is managed at the organisational level, U3 is an extension of U1. Where organisational units utilise specific external vendors (i.e. Data Processors), U3 is an extension of U2. In this, we consider the practical situations where data governance is often managed by internal units despite GDPR associating Data Processors directly with a Data Controller.

The key information flows between these stakeholders, in addition to U1 and U2, involve: (i) ROPA information from appointed processors; (ii) Provenance, e.g. sources, timestamps, contact details; (iii) Collation of information from heterogeneous sources into a common ROPA; and (iv) A selective part of ROPA, e.g. specific processor.

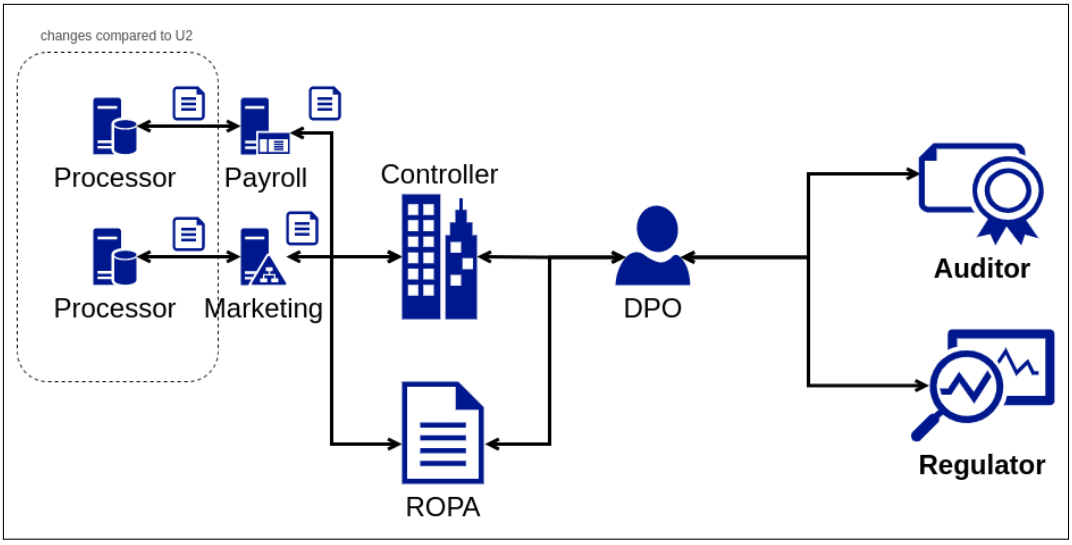

Figure 3. A Data Controller with Organisational Units and Data Processors

4.4. Use Case U4: Data Controller in a Joint Controllers relationship

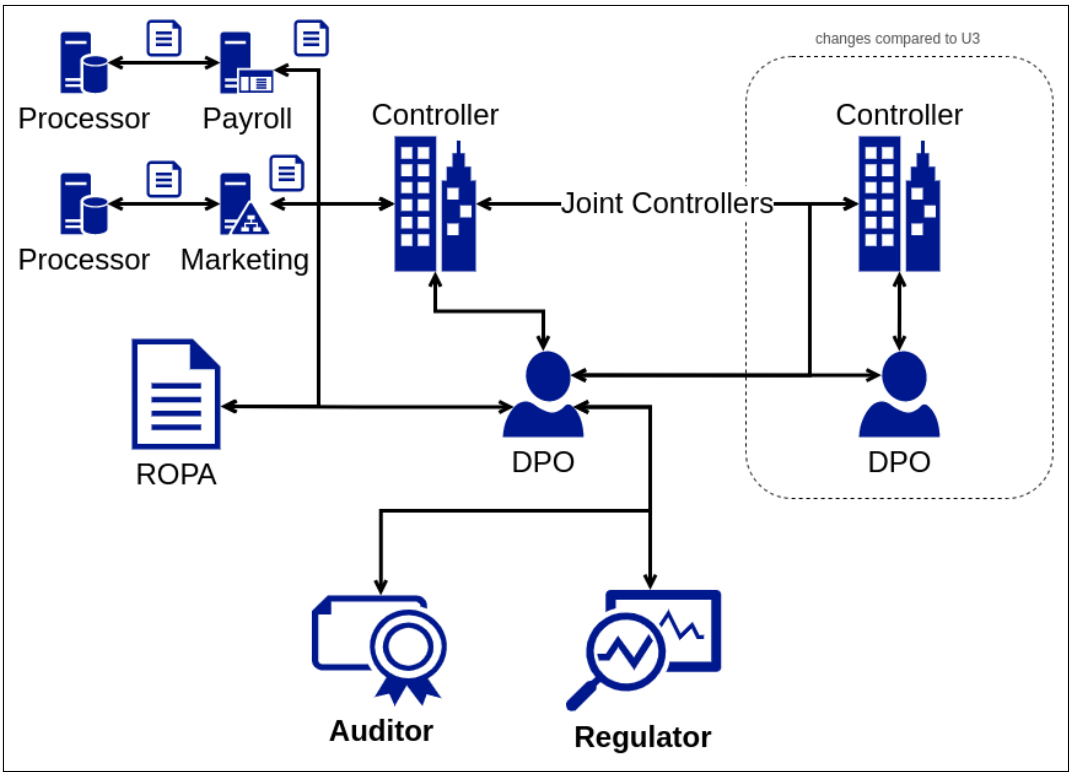

Figure 4. Data Controller in a Joint Controller relationship

The fourth use case, illustrated in Figure.4, expands U3 with the Data Controller being in a Joint Controllers relationship with two or more controllers sharing the responsibility of processing as per GDPR Art. 26. Similar to the possibility of associating processors with organisational units in U3, joint controllers can also similarly be associated with units for situations where the processing is limited to a unit's activities. In U4, the controller and its DPO have additional information flows regarding collecting relevant information from other (joint) controllers and any potential follow-ups.

The key information flows for these stakeholders, in addition to U3, involve: (i) ROPA information from joint controllers; (ii) Provenance, e.g. sources, timestamps, contact details;

(iii) Collation of information from heterogeneous sources into a common ROPA; and (iv) A selective part of ROPA, e.g. specific (external) controller.

4.5. Use Case U5: DPO overseeing multiple Data Controllers

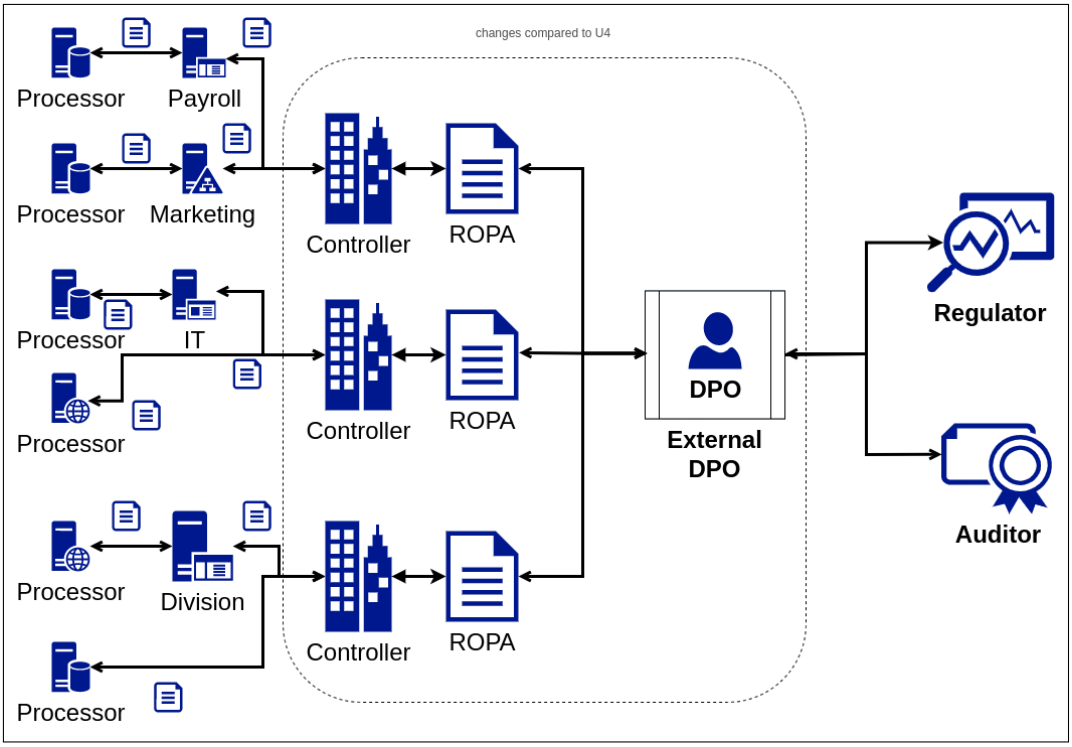

Figure 5. A DPO overseeing multiple Data Controllers

Use cases U1-U4 considered the perspective of a Data Controller that employs a DPO in terms of managing their ROPA information. In U5, illustrated in Figure.5, we consider the scenario of a DPO being an external organisation or individual providing 'DPO-as-a-service'. We call this entity 'External DPO', and consider their duties as involving overseeing multiple organisations. The external DPO has to address U1-U4 for several organisations which translates to additional information flows. This is distinct from information flows associated with other external entities, i.e. DPAs or auditors, in that the external DPO requires information including internal organisational units and data governance processes for an accurate understanding and potential follow-up tasks.

The key information flows for these stakeholders, in addition to U1-U4, involve: (i) Collect ROPA information from multiple organisations; (ii)Produce ROPA for a specific controller; (iii) Provenance, e.g. sources, timestamps, contact details; (iv) Separation of ROPA related information reflecting organisational units, e.g. departments; (v) A selective part of ROPA, e.g. specific department for a specific controller.

4.6. Requirements Analysis

Since the GDPR does not dictate or concern how a ROPA is generated or maintained, as long as it meets the legal requirements, the organisation has the freedom to determine practices that suit its compliance approach and style. For example, the organisation may choose to maintain ROPAs centrally overseen by the DPO, where information from all sources is fetched externally and collated into a common document (e.g. a spreadsheet) and added to the information management system. Alternatively, the organisation may opt to maintain separate ROPA documents for each of its departments.

In either case, upon being asked by a DPA or an auditor, the organisation has to produce a ROPA for the specified criteria, such as reflecting specific processing activities or for a certain time period. The organisation must first identify the relevant information from

its ROPA documents and extract the required information. This task can involve manual efforts by the DPO or responsible entity unless the organisation utilises technological solutions that support such use-cases and provide an easier workflow based on automation.

Identifying these distinct use-cases enables us to collect and harmonise the requirements regarding the expression of information and interpret them for the use of CSM-ROPA for data governance. We indicate that a solution must:

1. Indicate the source of information, e.g. department, processor
2. Represent collation of ROPA from discrete, possibly partial; information artefacts, e.g. purpose from the department, technical measures from processor
3. Record provenance, e.g. timestamps
4. Record organisational details, e.g. point of contact, responsible entity
5. Maintain distinct records, e.g. department or processors, or temporal periods

Further, we represent requirements for enabling systems utilising this information:

6. '*Packaging*': Sharing ROPA record(s) with internal or external stakeholders, e.g. department to DPO or processor to controller
7. *Querying*: Retrieving partial information from ROPA, e.g. specific period or process
8. '*Exporting*': Generating ROPA documentation as per requirements, e.g. GDPR Art.30
9. '*Customisation*': Customising information storage, retrieval, and exporting based on a variance in requirements, e.g. additional information for specific DPA templates
10. '*Assuring*': Providing data integrity and other quality guarantees for records

We also represent additional requirements that motivate operational details:

11. *Machine-readability*: for using automation and tooling for information management
12. *Interoperability*: for consistency in and interpretation across stakeholders
13. '*Openness*': for enabling adoption without lock-ins across technologies or providers
14. 'Extendability': to enable customisation of a solution for a use-case or contextual requirements, e.g. additional terms, new information requirements
15. 'Verifiable': to support information management through validation of information in terms of correctness and completeness, e.g. all necessary fields are declared with valid information types, as well as to support compliance processes in ensuring validity and accountability, e.g. ensuring every processing has a purpose

From this, we conclude that CSM-ROPA, while sufficient to represent information required to generate a ROPA, is insufficient to meet requirements for exchanging or using information amongst relevant stakeholders. In the next section, we present our use of this as a motivating factor for developing a 'catalogue' that can encapsulate the ROPA related information and satisfy requirements for its maintenance and exchange with stakeholders.

5. DPCat: A Data Processing Catalogue

In the earlier sections, we presented a 'Common Semantic Model for ROPA' (CSM-ROPA) representing a consolidated set of information requirements based on an analysis of DPA ROPA templates that can be used as a machine-readable and interoperable vocabulary through the use of DPV (see Section 3.). We then explored the sources and use of ROPA in terms of data flows between stakeholders and identified five use-cases that provided further requirements regarding using CSM-ROPA in practical settings. In this section, we present the Data Processing Catalogue (DPCat) specification, also published online[8], that addresses identified requirements and facilitates governance of information from intra- and inter-organisational heterogeneous sources to enable representation of ROPA in a machine-readable and interoperable manner.

DPCat extends the DCAT Application profile for data portals in Europe[9] (DCAT-AP) with concepts identified in CSM-ROPA using DPV to enable representation of ROPA and

[8] https://w3id.org/dpcat
[9] https://op.europa.eu/en/web/eu-vocabularies/dcat-ap

associated information as 'catalogues' and 'datasets' respectively, that can be recorded and exchanged between stakeholders. DCAT-AP is a profile of the Data Catalog Vocabulary[10] (DCAT v2) - a W3C standard for facilitating interoperability between data catalogues. DPCat maintains compatibility with DCAT-AP, and through it with DCAT, thereby enabling it to be used in all catalogue-based information management tools and data portals that support DCAT. In particular, the choice of DCAT-AP was made to present a mechanism for sharing ROPA related information using an EU-advocated standard and to promote the possibility of reusing existing data portal infrastructures for compliance-related purposes - such as requirements for ROPA between controllers, processors, and DPAs.

Our prior work [2,3] regarding DPCat was based on the CSM-ROPA developed from 6 DPA ROPA templates - which addressed 2 (U2: organisational units, U3: processors) of the 5 use-cases. This work incorporates: updated CSM-ROPA for 17 DPA ROPA templates, updates made to the DPV (from v0.2 to v0.5), and integration of DCAT and DCAT-AP requirements (e.g. cardinality) for compatibility.

5.1. DPCat Overview

DPCat, illustrated in Figure.6, distinguishes between ROPA (as a document or an artefact) and '*entries*' within a ROPA where each *entry* represents a specific context - such as a business process or data processing purpose. To represent these, we semantically extend the DCAT(-AP) concepts '*catalog*' and '*dataset*' as '*ROPA*' and '*ROPARecord*', respectively. We also extend '*catalog*' as '*ROPACatalog*' to represent a collection of *ROPA* catalogues (i.e. a catalogue of catalogues) for when an organisation has multiple ROPA documents e.g. representing different temporal periods or activities or organisational units (e.g. departments). Appendix C provides an overview of these concepts.

ROPARecord is a *dcat:Dataset* that catalogues information to be documented in a ROPA, is akin to a 'single row' in a ROPA spreadsheet and represents a single record of processing[11]. It is used as an instance of *dpv:PersonalDataHandling* to associate concepts such as purposes of processing or legal bases using the relevant DPV concepts identified from the CSM-ROPA analysis. To ensure compatibility with DCAT and DCAT-AP requirements and recommendations, such as a publisher being a *foaf:Agent*, DPCat declares the relevant DPV concepts as a subclass of DCAT(-AP) specified concepts.

A (*dpcat:*)*ROPA* represents a *dcat:Catalog* consisting of one or more *ROPARecord* datasets and reflects the conventional perspective of 'ROPA as a single document' with each entry being a *ROPARecord* within the catalogue. In both *ROPA* and *ROPARecord*, the DCAT properties provide an association with relevant information such as the publisher indicating who had produced or provided that record, temporal annotations such as when the record was produced, or the time period represented, and annotations such as titles and descriptions. A *ROPACatalog* is the same as a *ROPA* in terms of being extended from *dcat:Catalog* and is used to bundle related *ROPA* catalogues together using *dcat:catalog* relation.

For common ROPA-related communication between stakeholders, such as associating a 'point of contact' (e.g. department or manager) for that information, DPCat uses DCAT relation *dcat:contactPoint*. Additionally, to adhere to GDPR terminology, it uses the DPV properties to indicate controller (*dpv:hasDataController*), DPO (*dpv:hasDataProtectionOfficer*), and 'responsible entity' (*dpv:hasResponsibleEntity*). In this, the overlap between DCAT and DPV terms, such as the controller being the publisher or the DPO being the point of contact, may not always occur - such as when representing activities limited to a department where the point of contact is a member of that department who liaises with the DPO.

[10] DCAT Homepage, http://www.w3.org/TR/vocab-dcat-2/

[11] In a *ROPARecord* instance, the concepts are coherent i.e. all purposes apply to all personal data and are shared with all recipients and so on. To indicate separation, separate instances should be created.

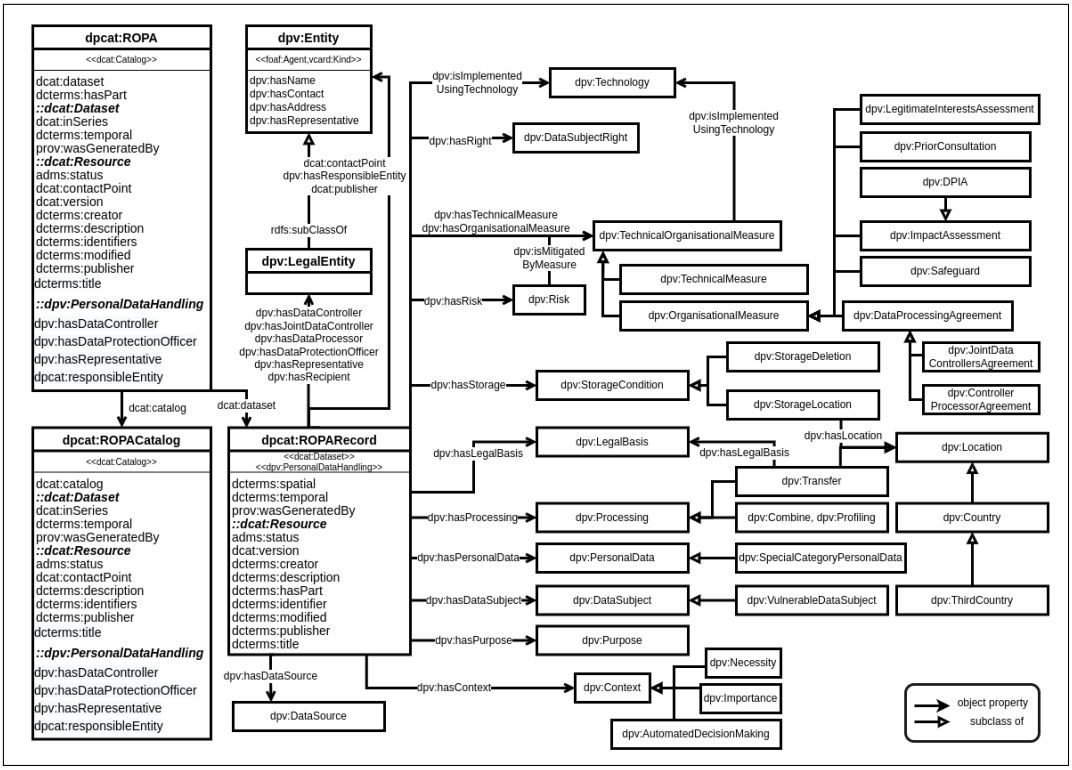

Figure 6. DPCat *ROPA*, *ROPARecord*, and *ROPACatalog* overview

5.2. Using DPCat for ROPA Information Management

As we elaborated on in Section X., the information and data governance requirements within the use-cases show a need for each entity to organise, maintain, and exchange relevant information to carry out ROPA related processes. DPCat, as a specification, supports automation through integration into tools used for information storage and retrieval (e.g. databases) and information management practices (e.g. documents and data catalogues). It can represent all ROPA related information or only catalogue metadata with links to the actual information stored externally (e.g. spreadsheets) as datasets. In either case, DPCat provides a consistent information structure that enables technological solutions such as querying, validation, and exporting (see next sections) to assist the relevant stakeholders in their tasks.

DPCat facilitates data governance for ROPA by incorporating the organisation's structural and managerial requirements. For U1, where a ROPA has to be maintained at the organisational level, the *ROPA* and *ROPARecord* data can be maintained centrally. For U2-U4, where there are heterogeneous sources of information, and it is desirable to record them in the same manner for provenance and follow-ups, the DCAT relations enable provenance of publishers and points of contact. In contrast, *ROPACatalog* enables collections of related information issued by, e.g. a department or a processor.

The semantics of DCAT provides flexibility in determining how ROPA information could be organised and stored without determining a single method or structure. For example, in addition to the structuring based on organisational units and external entities, it may be desirable to keep records based on contextual information - such as specific business processes related to a product or service. This can be achieved by creating additional *ROPACatalog* entries representing the other collection and linking them to relevant *ROPA* entries. Through this, organisations can achieve multifaceted approaches in using the ROPA information without data duplication.

The use of technological solutions advocated by DPCat faces a hurdle in that the sources of information in ROPA related workflows may not necessarily have the technical knowledge to produce consistent and valid metadata. For example, a DPO with the

necessary legal knowledge does not necessarily have or is concerned with the underlying technicalities of information storage and retrieval beyond what is necessary to perform their duties. In such cases, existing information storage and management mechanisms such as databases and spreadsheets can continue to be used by DPCat being integrated into them rather than acting as a replacement. For example, using a SQL database, the information represented in its tables would utilise the DPCat as a schema with the input provided through existing means, e.g., input forms or importing spreadsheets using controlled structures. Alternatively, using an RDF-based solution such as a triple-store, the forms or spreadsheets could be converted to DPCat by utilising mappings.

We envision DPCat to be integrated in to a typical workflow (i.e. U2-U4) for recording ROPA as follows. The source (e.g. department representative) generates a *ROPARecord* containing relevant information with provenance as the department. They use mechanisms available to them - e.g. a series of forms or a script that converts spreadsheets. This information is collated into a *ROPA* collection representing contextual grouping as determined by the organisational structure (e.g. maintained per department). For sources external to the organisation (e.g. a processor), the provided information is similarly stored in dedicated *ROPA* and *ROPARecord* entries and optionally integrated directly into relevant datasets (e.g. controller listing processor's technical measures in its ROPA). This can use technological solutions such as a database or a portal. To facilitate the structuring of ROPA records in an organised manner, *ROPACatalog* entries are used to collect and group *ROPA* entries according to some criteria, e.g. temporal period, legal counsel, or responsible managers.

5.3. Using DPCat for Querying and Validation

DPCat supports and enables a wide assortment of queries and validation approaches that utilise its metadata-based structure to perform information retrieval and verification tasks. DPCat can be a vital tool in technological solutions used for compliance-related processes through these. This section presents a few examples of queries and validation tasks that motivate the use of DPCat in an organisation's ROPA related processes.

A common query associated with ROPA is retrieving GDPR Art.30 information for a specific context, such as data transfers or covering some time period. DPCat supports such queries through DCAT and DPV metadata, e.g. indicating transfer locations as *dpv:DataTransfer* and *dpv:hasLocation*, and DCAT *dcat:temporalPeriod* to perform time-based filtering. An example of this expressed as a SPARQL query is provided in Listing.1.

```
?Entry a dpcat:ROPARecord .
?Entry dct:title ?title .
?Entry dct:publisher/dpv:hasName ?publisher .
OPTIONAL { ?Entry dcat:contactPoint/dpv:hasName ?contact } .
?Entry dct:created ?created .
?Entry dpv:hasProcessing ?transfer .
?transfer a dpv:Transfer .
# minimum date within which data transfer occurs
?Entry dct:temporal/time:hasBeginning/time:inXSDDate ?start .
FILTER (?start < "2021-01-01"^^xsd:date) .
OPTIONAL {
    # maximum date, if available
    ?Entry dct:temporal/time:hasEnd/time:inXSDDate ?end .
    FILTER (?end > "2022-01-01"^^xsd:date) .
}
OPTIONAL { ?transfer dpv:hasDataImporter/dpv:hasName ?importer . }
OPTIONAL { ?transfer dpv:hasDataExporter/dpv:hasName ?exporter . }
```

Listing 1: SPARQL query retrieving ROPA records involving data transfers in a time period

Similar to querying, DPCat also supports verification and validation of information, typically ensuring or assessing compliance with the GDPR. Validation refers to whether sufficient information is available, is in the correct form and format, and is sufficient according to some requirements. Verification refers to the evaluation of the information based on some norms, such as specific obligations of the GDPR.

Constraints based on mandatory fields as prescribed by DCAT and DCAT-AP specifications also apply to DPCat since it extends them. Therefore, data represented using DPCat can utilise existing validation and verification mechanisms for conformance to these standards. In addition, DPCat promotes the expression of GDPR-specific constraints that are typically expressed as guidelines by DPAs and have been the subject of research by academic and commercial offerings. However, DPCat has an advantage over these existing solutions in that it also promotes interoperability between such verification mechanisms by virtue of being an interoperable specification for information to be verified.

As an example of information validation typically involved for GDPR, Listing.2 presents a SHACL constraint that ensures every *ROPARecord* instance has an associated purpose. In addition to ensure information is present and in correct form, SHACL constraints are also useful towards GDPR compliance, such as for ensuring an appropriate legal basis[12] as follows: (i) It must have a corresponding legal basis from GDPR Art.6; (ii) If processing involves special categories of personal data, it must additionally have a corresponding legal basis from GDPR Art. 9; (iii) If processing involves data transfers to non-EU locations, it must additionally have a corresponding legal basis from GDPR Art.45 or Art.46 or Art. 49. We plan to provide such SHACL shapes for both information validation and GDPR-based requirements verification in the future.

```
dpcat:Shape_EnsurePurpose
    a sh:NodeShape ;
    sh:name "Ensure every processing has a denoted Purpose "@en ;
    sh:description "Ensure the dpv:hasPurpose property is defined and has a
    value that is an instance of dpv:Purpose"@en ;
    sh:targetClass dpcat:ROPARecord ;
    sh:property [
        sh:path dpv:hasPurpose ;
        sh:class dpv:Purpose ;
        sh:minCount 1 ;
    ] .
```

Listing 2: SHACL constraint to ensure every *ROPARecord* has an associated Purpose

5.4. DPCat for Interoperable Information Exchange

DPCat provides a machine-readable and interoperable representation of information that an organisation can use to automate its ROPA management and associated tasks. In cases where information has heterogeneous sources, especially when involving external stakeholders such as processors and other controllers, DPCat can be utilised as a 'standardised information representation' for convenience in information flows. In this section, we explore the potential for such developments.

When they hire Data Processors, a Data Controller's obligation includes maintaining information about the processing activities outsourced to the processor and some specifics regarding how they are carried out. For example, controllers may ask processors to provide the technical and organisational measures they implement to ensure sufficient safety and security in processing. Similarly, controllers may require information for data storage locations of data for cross border data transfers. In cases where a processor contracts another (sub-)processor to carry out the processing, it has to maintain similar records of the sub-processors operations, but it also provides them to the controller as requested. In all these, information has to be periodical - maintained independently by the entity itself, communicated to other entities as contextually necessary, and the other entities also maintain this information independently. Such information flows and requirements are also necessary for a joint controller's relationships regarding involved controller(s).

If two entities communicating information for ROPA related tasks use DPCat for their internal information representation, they can directly exchange ROPA information using DPCat specified records. This is an ideal scenario. However, even if either or no

[12] Though GDPR Art.30 does not require a legal basis in a ROPA, DPA guidelines strongly recommend it.

entities do not use DPCat internally, DPCat can be utilised as a common specification for exchanging ROPA information between entities. In this case, the sender entity converts whatever internal representation it has into DPCat and sends it to the receiver entity to ensure that it can understand and interpret the information. The receiver converts DPCat based information to whatever internal representation they utilise. Thus, DPCat offers advantages for ROPA information exchanges even if organisations do not wish to adopt it completely for internal processes. DPCat is also useful for DPAs and auditors in the same manner where they can utilise it as an interoperable format for requesting information from organisations. The consistency and machine-readability of DPCat provide investigators with the potential for using automation and tools to reduce workload and repetitions.

6. Demonstration of DPCat in a Real-World Use-Case

In this section, we demonstrate the application of DPCat in representing real-world ROPA documents published by the European Data Protection Supervisor[13] (EDPS), perform validations of it using SHACL, retrieve relevant information using SPARQL queries, and export it as RDF graphs as well as spreadsheets adhering to DPA templates. We provide evidence for the practicality and feasibility of DPCat and its benefits in ROPA information management processes. The data, code, and outputs are available online[14].

6.1. Representing Information Using DPCat

EDPS is the DPA responsible for overseeing compliance by EU institutions, which consists of many employees across the various EU bodies and their associated personal data processing activities. The EDPS has published detailed ROPA documents based on GDPR Art.30 requirements that provide transparency and accountability. As of March 2022, the EDPS has made available 58 ROPA document collections - with each consisting of one more PDF (format) document providing information in English regarding the processing operations. Collections are structured based on 'topics' - which can be a department (e.g. Administrative and Human Resources, or IT), processes (e.g. Communication, or Public Events), or specific measures (e.g. Access to documents, or Physical Security).

We analysed EDPS ROPA documents and selected four (ids: *01, 05, 13, 55*) that covered the U1-U4 use-cases for departments, processors, joint controllers, and data transfers. We did not include the other documents despite their relevance due to the large labour and analysis efforts required, and because the selected documents sufficed in demonstrating DPCat's application. The documents were PDFs, intended for human comprehension, and lacked consistent semantics - e.g. *purpose* field also contained *legal basis*.

We interpreted these documents and their structure as follows: each document (i.e. PDF) represented a single *ROPA* instance, and the information contained within them structured using *ROPARecord* instances. We utilised the criteria that each *ROPARecord* would adhere to a single 'contextual entry' based on qualitative criteria regarding the complexity of information and separation of concerns. For example, document X specified two processors, which we interpreted as separate *ROPARecord* instances for each processor to indicate the separation of concern in the controller's communication and data governance. The entire collection of documents and RDF graphs were then expressed as part of a single *ROPACatalog* instance reflecting the published set of records on EDPS' website.

The manually created RDF graphs were enhanced using the Apache Jena RDFS reasoner[15] to create a 'complete graph' for simplifying querying and validation. The limited RDFS reasoning was sufficient here to obtain the expansion of subclasses and subproperties within the graph rather than generating inferences using an OWL reasoner. For storing the information and offering a querying interface, we utilised GraphDB[16] Free Edition triple store, as it is a freely available triple-store compliant with relevant standards (e.g. SPARQL)

[13] https://edps.europa.eu/about/data-protection-within-edps/records-register_en
[14] https://w3id.org/dpcat/demo/edps-ropa
[15] https://jena.apache.orgl
[16] https://www.ontotext.com/products/graphdb/

and has several features for convenience, e.g. friendly interface, integrated reasoners, SHACL validation. An overview of data workflow is provided in Figure.7:

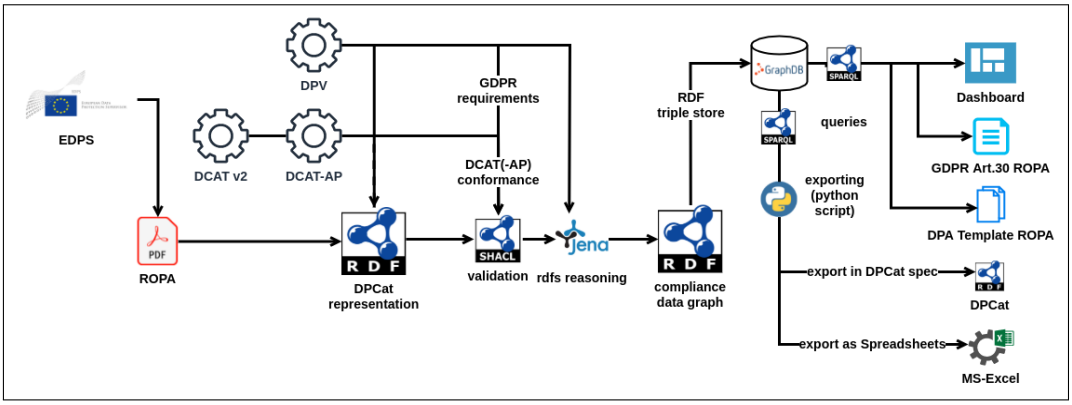

Figure 7. Data workflows in DPCat demo application

6.2. SHACL Shapes for DPV

For verification and validation of the generated RDF graphs, we first utilised the SHACL constraints provided with DCAT-AP specification to ensure data correctness according to DCAT and DCAT-AP defined requirements, e.g. publishers being of type *foaf:Agent*. We then developed and utilised SHACL shapes representing the cardinality and type constraints to ensure correctness for DPCat's requirements. For executing the constraints, we utilised the open-source and freely available TopBraid SHACL[17] tool.

In performing validation of the information, the shape constraints are based on DPCat, which utilises DPV and DCAT concepts to represent relevant information. However, neither DPV nor DPCat indicates what 'shape' some information must be represented. Consequently, there may be more than one 'shape' for a given scenario, often at arbitrary levels of complexity, which prevent a single set of common SHACL shapes from being developed and provided alongside the DPCat specification. For example, a SHACL constraint for ensuring data transfers are specified along with their appropriate locations can be modelled in terms of *dpv:hasLocation* of *dpv:DataTransfer*. However, the *DataTransfer* instances could be used at any arbitrary node within the graph, making it difficult to define follow-up constraints such as the recipient of that transfer and its location.

A simple solution would be to associate all the relevant fields with a *ROPA* or *ROPARecord* instance. A challenge in this is that all the DCAT-based structure may not be capable of incorporating all fields or that it would make DPCat too complex. An alternate approach would be identifying use-cases for each concept's use and defining specific SHACL shapes for how that information should be expressed using DPV. Given that this requires significant analyses and effort, for the purposes of this article, we limited our defined SHACL shapes for representing information from the EDPS documents. However, we argue for further research and development of such shapes so that they can be used to ensure data is consistently represented across use-cases and implementations.

6.3. Querying ROPA Information

To simulate typical tasks performed by a DPO or a DPA, we utilised SPARQL queries for two use-cases: (i) retrieval of information required by GDPR Art. 30; and (ii) overview of practices within an organisation in terms of various organisational units, purposes, legal bases, recipients, data transfers, etc. Here, query (i) relates to common compliance documentation procedures, and query (ii) shows the potential for DPCat to help create internal reports or dashboards based on ROPA information, e.g. for a DPO.

[17] https://github.com/TopQuadrant/shacl

The first query, shown in Listing.3 with an output snippet in Table.1, retrieves ROPA information as per GDPR's Art. 30.

```
SELECT DISTINCT ?Entry ?title ?purpose ?datasubject ?personaldata
    ?recipient ?legalbasis ?transfer_location
WHERE {
    ?Entry a dpcat:ROPARecord .
    ?Entry dct:title ?title .
    ?Entry dpv:hasPurpose/skos:prefLabel ?purpose .
    ?Entry dpv:hasDataSubject/skos:prefLabel ?datasubject .
    ?Entry dpv:hasPersonalData/skos:prefLabel ?personaldata .
    OPTIONAL { ?Entry dpv:hasRecipient/dpv:hasName ?recipient } .
    OPTIONAL { ?Entry dpv:hasLegalBasis/skos:prefLabel ?legalbasis . }
    OPTIONAL { ?Entry dpv:hasProcessing ?processing .
        ?processing a dpv:Transfer .
        ?processing dpv:hasLocation/skos:prefLabel ?transfer_location } }
```

Listing 3: Obtaining a GDPR Art.30 based overview using SPARQL

Table 1. Query results (summarised) for GDPR Art.30 information using DPCat

Title	Purpose	Data Subject	Personal Data	Recipient	Legal Basis	Transfers
Selection of staff	Staff Selection	Job Applicants	Applicant CV	Selection Panel	Staff Reg. 2020	
Financial Transactions	Payment	Staff members	Physical Address			
Financial Transactions	Payment	Staff members	Credit Worthiness	AirPlus		Third Country
Financial Transactions	Budgetary commitments	Staff members	Job Applicant CV	ERCEA's Speed-well operators		
Financial Transactions	Budgetary commitments	Staff members	Bank Account	Local Profile Manager		
Financial Transactions	Payments	Staff members	Bank Account	The EDPS Financial team		

The second query, shown in Listing.4 with an output snippet in Table.2, provides an overview of the organisation's processing activities and relationships with external entities by retrieving relevant information from *ROPARecord* instances.

```
SELECT DISTINCT ?org ?title ?purpose ?processor ?jointcontroller
WHERE {
  ?record a dpcat:ROPARecord ; dct:title ?title .
    ?record dct:publisher/dpv:hasName ?org .
    ?record dpv:hasPurpose/skos:prefLabel ?purpose .
    OPTIONAL { ?record dpv:hasDataProcessor/dpv:hasName ?processor }
    OPTIONAL {
        ?record dpv:hasJointDataControllers/dpv:hasName ?jointcontroller } }
```

Listing 4: SPARQL Query for overview based on GDPR Art.30 using DPCat

6.4. Exporting a ROPA

To demonstrate how DPCat can facilitate information exchange and data governance within and between stakeholders, we provide two examples of information being exported. The first example exports information as DPCat defined catalogues by using SPARQL CONSTRUCT queries to retrieve related information as an RDF graph. The SPARQL query and the resulting graph can be viewed online. Such exports help store information in the form of backups or copies or create documentation in graphs. It is also helpful in exchanging ROPA information between stakeholders, such as those accompanying data governance between Data Controllers and Data Processors, that all support use of DPCat.

The second example simulates automation of a DPO manually managing information in a spreadsheet. For this, we utilised a Python script that executed SPARQL queries and exported results into an MS-Excel (.xlsx) document based on DPA ROPA templates. While the output of a SPARQL query itself could also be exported as a CSV document, the use of Python in this case was to replicate the structure and contents of the DPA template and to operate over the more complex XLSX format that supports tabs within spreadsheets.

Table 2. Query results (summarised) for GDPR Art.30 information using DPCat

Department	Process	Purpose	Data Processor	Joint Controller
Human Resources, Budget, Administration (HRBA) Unit	Staff Selection	Select staff for the EDPS and EDPB Secretariat		
Human Resources, Budget, Administration (HRBA) Unit	Selection and management of interim staff	Monitoring of 7-year rule (EDPS Decision 13.12.2018)		
Human Resources, Budget, Administration (HRBA) Unit	Communicate staff selection	Select staff for the EDPS and EDPB Secretariat	Randstad Belgium SA/NV	
Human Resources, Budget, Administration (HRBA) Unit	Communicate staff selection	Select staff for the EDPS and EDPB Secretariat	Daoust SA/NV	
Human Resources, Budget, Administration (HRBA) Unit	Payment of Invoices for services	Payment of invoices for services		
Human Resources, Budget, Administration (HRBA) Unit	Communicate staff selection	Communicate staff selection	Randstad Belgium SA/NV	
Human Resources, Budget, Administration (HRBA) Unit	Communicate staff selection	Communicate staff selection	Daoust SA/NV	
Human Resources, Budget, Administration (HRBA) Unit	Administration of Access Requests	Administration of Access Requests		
Human Resources, Budget, Administration (HRBA) Unit	Financial Transactions	Financial Transactions	EC - DG-BUDG	EC - DG-BUDG

6.5. Analysis of Implementation and Lessons Learned

The application of DPCat to real world ROPAs exposed inherent difficulties in constructing semantic representations due to inputs lacking or being loosely structured as opposed to strict structure machine-based tools require. We discussed exploring this issue further with a proposed solution where a separate *registry of controlled vocabularies* is created by the organisation for use-case to first register their concepts, such as the specific purpose used, or data category processed, and to then ensure the ROPA documents only used these concepts. However, we found this solution to significantly deviate from the organisational processes that lack such structured data collection methods. We consider this an open problem with the hope of better tooling being able to resolve it.

In representing the ROPA information using DPV, we faced hurdles in that the DPV as a vocabulary can support a wide range of data modelling styles. This presents barriers to the use of DPCat as a common information representation mechanism as two different organisations can model their data differently. While the common conceptual structure of DPV can assist in aligning the two models, it is better for the development of tools to have a consistent information structure. For this, we propose the creation of '*DPV Shapes*' that provide suggested data modelling practices for modular use-cases. Such shapes, expressed using SHACL, will foster commonality in how the DPV is used, and will act as a common model for other modelling approaches can be reduced or aligned to. In this, it is important to state one of the strengths of the DPV is its lack of rigid adoption requirements, which provides an adopter the flexibility to use it within their use-cases. The provision of shapes enables continued flexibility of the DPV as a vocabulary while providing guidelines for how it can be consistently used or made interoperable across different applications.

Lastly, we faced challenges in determining a suitable mechanism for validation of DPCat specified information. While we utilised SHACL shapes to demonstrate the potential for such validations based on information and GDPR compliance requirements, this area merits further exploration. In particular, SHACL constraints can be used for two categories of evaluation: first to check whether the necessary information is present and has expected values - similar to DCAT-AP SHACL shapes. The second is based on requirements drawn from the GDPR, such as ensuring the correct legal basis is used. In these, the first is an inherent evaluation of *conformance* to a specification - as seen from the cardinality constraints in DCAT and DCAT-AP, while the second directly addresses GDPR compliance verification. This follows earlier research explorations demonstrating use of SHACL constraints in ensuring information correctness and conformance for GDPR compliance [34].

7. Discussion

7.1. Impacts on ROPA Related Research

Our literature review (Section.2) shows that although GDPR compliance is well studied, there has been a lack of academic research specifically addressing ROPAs. Labadie and Legner [10] identify "maintenance of records of processing activities" as a core GDPR data management (sub-)capability for organisations. DPCat transforms this into an IT system capability by extending Labadie and Legner's model. In addition, DPCat adds sub-capabilities for aggregation of diverse accountability data, exchange of machine-readable ROPA information with stakeholders, generation of DPA-specific compliance records, and assurance mechanisms for ROPA data quality (Figure AB).

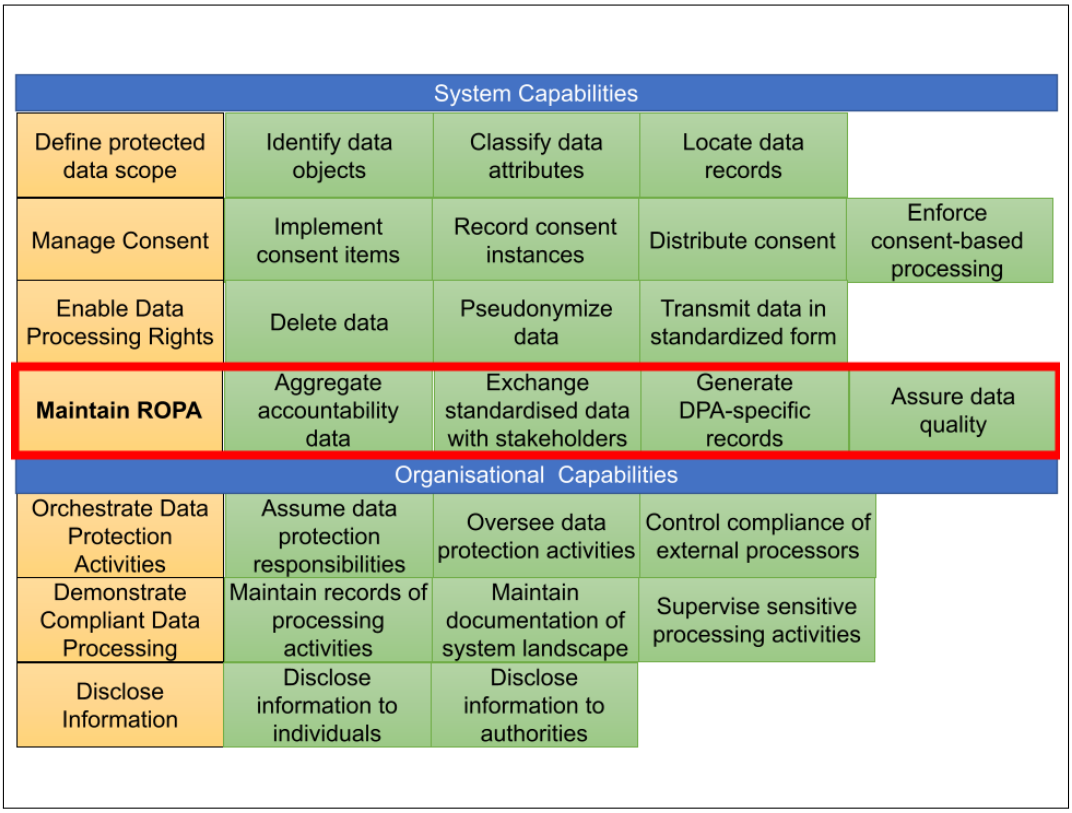

Figure 8. Labadie & Legner Capability Model for Data Management in GDPR [10] extended with "Maintain ROPA" system capabilities using DPCat - highlighted with bold text and red borders

In addition, the analysis of DPA ROPA templates (Section.3) has demonstrated the previously undocumented extent of variance between DPA approaches, which impacts DPOs and GDPR-aware system designers or integrators. The creation of CSM-ROPA provides an ontological structure for representing ROPA related information. By utilising DPV as a community-endorsed specification that draws on the skills, requirements, and expertise of DPOs, legal experts, and technologists - it provides a strong basis for establishing an agreement of semantics to address the gap identified by existing efforts [10,21,22]. In addition, CSM-ROPA supports existing approaches based on semantics by providing a target ontology, such as for Enterprise Architecture models proposed by Huth [12].

By moving ROPA processes to a data cataloguing approach, the DPCat specification facilitates adoption of modern metadata-driven data governance [33]. This in turn motivates adoption of data stewardship to support intra- and inter-organisational heterogeneous sources of information and compliance [10]. It also aligns with recent EU recommendations on placing data governance at the centre of personal and AI-based data processing [35]. These advances are significant when compared with low levels of governance and automa-

tion of ROPAs exhibited by most organisations to date and provide a step towards effective technologies and tools for data protection.

7.2. Impacts on Real World GDPR Compliance

This section discusses impacts of DPCat on ROPA governance: enforcing consistency on current largely manual exercises for ROPA data maintenance, impact on privacy and data protection software, and impact on data exchange within large organisations.

The application of DPCat to EDPS ROPAs highlighted the cumbersome efforts required in disentangling machine-readable data from human-oriented documentation, even when data is in semi-structured tabular form. The human creators of ROPA information tried to enforce consistency and ease of presentation through such structures, but the variations and inconsistencies in readily using these as machine-readable information were apparent when converting them to RDF using DPCat. In addition, cross-references or inferences had to be made from other PDF documents in the dataset, leading to difficulty of validation and more opportunities for inconsistency. By providing a flexible, layered approach to machine-readable ROPA information collection, DPCat supports analysis and progressive integration or automation of processes from lightweight metadata-oriented approaches that emphasise DCAT-AP and GDPR Article 30 requirements to more detailed knowledge models of accountability information that expand into full DPV models of personal data handling and beyond to other ontologies or datasets. No matter what level of modelling is selected, DPCat will provide significant advantages in terms of the ease and consistency of ROPA data maintenance in an organisation and, thus, the cost of compliance.

The commercial software offerings for data protection [7] consist of tools supporting organisational units that are primarily aimed at DPOs or compliance units which maintain documentation for the organisation. However, they are information silos as they lack interoperability with other systems. In response, software companies may develop APIs or adapters to connect these to other systems, which will require updates and integration efforts for every system used by all stakeholders. This becomes particularly problematic when organisations appoint new processors, acquire companies, or add new systems. In contrast, DPCat provides a single integration point for ROPA information that can be used by any data protection, compliance, and operational business systems.

DPCat also addresses the information needs of real-world intra-organisation use cases that are absent from existing literature (U1). For example, one co-author is a DPO for an organisation consisting of five divisions, with their own DPOs, containing 29 affiliated legal entities spread across UK, EU, and USA. As would be the case in such organisations, there is a large amount of intra-group information flows. The organisation utilises shared services such as IT support, data analytics services, and human resources. There are also a large number of appointed processors providing services to affiliates. The organisation has a limited number of joint controller relationships, and intra-organisation processing is more common. Here, the challenge for the DPO [13] is ensuring the complexity of processing activities is collected and accurately reflected in the numerous ROPA documents. For this, the organisation uses several standalone proprietary solutions that still create a large dependence on manual effort and documents. Applying the DPCat approach to ROPA data management in such a landscape enables greater automation and spans the heterogeneous IT systems involved, both for compliance and business operations. Coping with an ever-changing diversity of internal data processing links is the key to empowering the DPO to monitor the personal data processing, communicate with stakeholders, and identify non-compliance. This then facilitates better management of external relations and compliance activities governed by contracts and law.

7.3. Practical Challenges for DPCat Deployment

Requirement for Enhanced Data Governance: Despite many organisations embracing the productivity and agility gains of digitalisation, they continue to struggle with the basic principles of data governance [9]. The agreed uses that data is put to must be clearly

defined, and the organisation must ensure that the use of data positively relates to the regulatory environment. Organisations need to define the agreed behaviours and policies for data quality, who will access the data, how data is interpreted, and how long the data will be retained. The challenge organisations face regarding personal data is locating, classifying, and cataloguing accountability data. DPCat provides an incentive to deploy a machine-readable data catalogue platform such as CKAN[18] that provides user-friendly interfaces and tools supporting all these activities. In addition, the presence of an actively maintained data catalogue is a spur to wider data governance activities in the organisation.

When examining the federated or distributed aspects of ROPA data governance deployment, the role of Processors is significant. They will need to contribute updates to ROPA as per the DPCat specification. There may be resistance to this activity, or the Processor may need technical assistance to meet the DPCat requirements. A wide adoption of DPCat by stakeholders would bring great benefit, but even if one company chooses to use it , it brings benefit, as DPCat can be utilised as an export format. DPCat provides some assistance here by being a single comprehensive integration point; if Processors can comply with it, then unlike proprietary solutions, this integration cost should be a reusable effort that can be used for many customers playing the role of the data controller.

Agreed Semantics: The need for an agreed ontology to describe data processing activities is key to DPCat's success.This lack of common understanding of privacy terms is limiting the growth of the privacy tech industry [21]. DPCat provides a solution that vendors could adopt since it is based on two existing standards, DCAT-AP and DPV. Nonetheless, vendors are typically driven by their customers or regulators to adopt open standards and hence the importance of the role of DPAs as discussed in the next section. Unless DPAs get involved, it will be up to DPOs and other privacy software customers to demand interoperable solutions for ROPA management.

Role of DPAs: DPAs could have a significant role in automated regulation, making compliance easier to achieve [36]. When we compare the success of RegTech, we see regulators that enable and facilitate digital compliance, actively promote and enable digital regulatory compliance standards, and act as enablers for the automation of regulation to actively create an environment for digital compliance [9]. Adopting DPCat for Data Controller to DPA communication would benefit the DPA when auditing ROPAs as audit, a breach investigation or inspections. In order to achieve the adoption of DPCat, DPAs would need to move towards a symbiotic relationship with technology innovators and organisations that process personal data and develop open-source compliance tools, digital regulations, and sandboxes [37] as well as agreed common semantic vocabularies like DPCat [9]. A proactive DPA could certainly speed up the use of DPCat, and there are certainly some moves toward technology, such as an online DPIA template [38]. However, we are very much at the early stages of automated GDPR regulatory compliance.

7.4. Limitations

While the approaches motivated by CSM-ROPA and DPCat provide promising solutions to the challenges in data governance associated with maintenance and use of ROPA towards GDPR compliance requirements, it also has certain limitations that need to be addressed to ensure it is effective in practice. In this section, we discuss identified limitations and propose future efforts toward addressing them.

Limitations of Scope: The DPCat specification reflects the information requirements derived from CSM-ROPA, which was constructed based on the GDPR requirements, and DPA guidelines and templates regarding ROPA. While this makes DPCat sufficient to carry out tasks associated with ROPA, it does not consider the relevance and overlap of information between a ROPA and other compliance documents - such as DPIA (Data Protection Impact Assessment), TIA (Transfer Impact Assessment), Data Breach records, and Controller-Processor or Controller-Controller agreements.

[18] https://ckan.org/

In each of these, there is an obvious overlap with some of the information stored in a ROPA and the necessity to link these to the ROPA itself. For example, a DPIA may concern several processing activities that are spread across distinct *ROPARecord* instances. While the ROPA can link to the DPIA document trivially through single information, it is advantageous for the DPIA information to be expressed similarly as the ROPA information so as to enable better information interoperability and governance and motivate the creation of tools that can work on all compliance based activities using the same information. This can be achieved by further developing DCAT-based solutions for all of the information necessary in legal compliance tasks - for the above mentioned requirements.

Limitations of Vocabulary: The DPV forms an important aspect of DPCat in that it provides the vocabulary for representing GDPR-associated terms in a machine-readable and interoperable manner. Therefore, any limitations of DPV will also be reflected within the capabilities of DPCat. Given that the DPV is a community-managed resource (through the DPVCG), there is a forum for proposing additions and enrichments as needed for DPCat's applications. However, better alignment between DPCat and DPV versions will have to be established so as to provide the reliability of DPCat's usage and interpretation - for example by pinning DPCat's use of DPV to a specific version.

Limitations of Jurisdiction: DPCat as a solution is EU-centric in that it directly addresses (only) GDPR requirements. However, there may be a wider need for organisations to document their processing activities in a different jurisdiction or a jurisdiction-agnostic manner. For addressing cases, DPCat may be supplemented with extensive modifications, such as an adopter's own jurisdiction-specific vocabulary, which may bring about incompatibility between implementations. A solution would be to develop DPCat into a domain and jurisdiction agnostic specification and then provide the GDPR specific concepts as an extension of the profile. This reflects current work regarding extending DCAT to DCAT-AP, and the provision of GDPR-specific concepts[19] separate from the 'main' DPV.

Limitations of 'Data Shapes': As mentioned earlier in Section 6., the querying and validation of information require consistency or foreknowledge regarding how the data is structured or 'shaped'. Without this, the resulting SPARQL queries and SHACL shapes can be difficult to express or become complex without this. To ensure the consistency of DPCat implementation, especially for information exchange, it is vital to enforce the consistency of the underlying information. While DCAT (and DCAT-AP) provide this consistency to the expression of catalogues and datasets as resources, the lack of such consistency in the expression of DPV specified information needs to be addressed. For this, we propose the development of use-cases that define expectations and requirements for information, e.g. a data transfer must specify location", to create corresponding 'SHACL shapes' that harmonise how different implementers should utilise DPCat specified information. This activity can be undertaken by the DPVCG for the larger benefit of all DPV adopters or, failing that, within DPCat to ensure its consistency in application.

8. Conclusion

The heterogeneity of data sources representing the organisation's data processing activities, presents significant challenges when completing a ROPA. Our research sought to establish the extent to which the DPCat specification for an interoperable and machine-readable data processing catalogue based on DCAT-AP and DPV could overcome the heterogeneity of sources to facilitate the preparation of a ROPA.

We have shown that the DPCat specification enables more automation for realistic distributed ROPA maintenance use cases, leading to stronger regulatory compliance. DPA's use of DPCat could also ease the investigative burden required for effective enforcement. In pursuit of our first research objective (*RO1*) to identify the information necessary to represent ROPAs, we reviewed 17 ROPA templates across 31 DPAs. Our analysis identified 47 unique GDPR concepts, with templates requiring a minimum of 18 concepts up to a

[19] https://w3id.org/dpv/dpv-gdpr

maximum of 32 concepts. Over the past three years, the DPV has been enhanced to express these concepts, and currently, 44 of the 47 concepts can be expressed exactly, and one can be partially expressed. The two remaining concepts are with the DPVCG for consideration.

For the second research objective (*RO2*), we presented the Data Processing Catalogue (DPCat) specification that facilitates governance and maintenance of data from intra- and inter-organisational heterogeneous sources to enable representation of information related to ROPA. Its application to EDPS ROPA demonstrated how DPCat could be utilised as a machine-readable solution to overcome conventional limitations for when data is maintained in documents or proprietary systems lack machine-readability and interoperability.

The EDPS application also showed how DPCat enabled a data controller/processor to describe processing activities using a standardised model and vocabulary that facilitated aggregation, querying, validation, and exporting from heterogeneous sources (*RO3*). We used SHACL to ensure correctness, and SPARQL to query and export information for GDPR articles and DPA templates. Through this, we establish the data quality governance process for ROPA by harmonising inputs from heterogeneous sources and producing dynamic documentation that accommodates differences in regulatory approaches across DPAs.

In addition to formulating a research problem, we also explored the potential impact in real-world situations through the use-case, application, and discussions, and identification of concrete future directions to ensure practical benefits from implementing our work. In addition, as DPCat is an interoperable machine-readable record of the personal data processing activities of organisations, it offers avenues of future research, such as the generation of privacy notices, DPIAs, automatic supplier due diligence checking and international transfer compliance assessments from a common information model.

Funding: This research has received funding from Uniphar PLC, and the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106_P2) and co-funded by the European Regional Development Fund. Harshvardhan J. Pandit has received funding under the Irish Research Council's Government of Ireland Postdoctoral Fellowship Grant#GOIPD/2020/790.

References

1. Registers Of Processing Activities - Castlebridge. https://castlebridge.ie/research/2020/ropa-report/.
2. Ryan, P.; Pandit, H.J.; Brennan, R. A Common Semantic Model of the GDPR Register of Processing Activities. In Proceedings of the Frontiers in Artificial Intelligence and Applications; Villata, S.; Harašta, J.; Křemen, P., Eds. IOS Press, 2020. doi:10.3233/faia200876.
3. Ryan, P.; Brennan, R. Demonstrating GDPR Accountability with CSM-ROPA: Extensions to the Data Privacy Vocabulary. In Proceedings of the Ryan, Paul and Brennan, Rob ORCID: 0000-0001-8236-362X <https://orcid.org/0000-0001-8236-362X> (2021) Demonstrating GDPR Accountability with CSM-ROPA: Extensions to the Data Privacy Vocabulary. In: 24th International Conference Enterprise Information Systems (ICEIS '21), 26-28 Apr 2021, Online.; ICEIS: Online, 2021; pp. –.
4. Records of Processing and Lawful Basis - ICO. https://ico.org.uk/for-organisations/accountability-framework/records-of-processing-and-lawful-basis/, 2022.
5. Ryan, P.; Pandit, H.; Brennan, R. Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-Related Information for GDPR Using DCAT-AP and DPV. In Proceedings of the Further with Knowledge Graphs. IOS Press, 2021, pp. 169–182. doi:10.3233/ssw210043.
6. Pandit, H.J.; Debruyne, C.; O'Sullivan, D.; Lewis, D. An Exploration of Data Interoperability for GDPR. *International Journal of Standardization Research (IJSR)* **2018**, *16*, 1–21. doi:10.4018/IJSR.2018010101.
7. IAPP Tech Vendor Report 2021. https://iapp.org/media/pdf/resource_center/2021TechVendorReport.pd
8. OneTrust. IDC Releases First Worldwide Data Privacy Management Software Market Shares Report - OneTrust. https://www.onetrust.com/blog/idc-releases-first-worldwide-data-privacy-management-software-market-shares-report/.

9. Butler, T.; O'Brien, L. Understanding RegTech for Digital Regulatory Compliance. In *Disrupting Finance*; Lynn, T.; Mooney, J.G.; Rosati, P.; Cummins, M., Eds.; Springer International Publishing: Cham, 2019; pp. 85–102. doi:10.1007/978-3-030-02330-0_6.
10. Labadie, C.; Legner, C. Understanding Data Protection Regulations from a Data Management Perspective: A Capability-Based Approach to EU-GDPR. In Proceedings of the Wirtschaftsinformatik, 2019, p. 15.
11. Martínez González, M.M.; Alvite Díez, M.L.; Casanovas, P.; Casellas, N.; Sanz, D.; Aparicio de la Fuente, A. OntoROPA Deliverable 1. State of the Art and Ambition., 2021.
12. Huth, D.; Tanakol, A.; Matthes, F. Using Enterprise Architecture Models for Creating the Record of Processing Activities (Art. 30 GDPR). In Proceedings of the 2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC), 2019, pp. 98–104. doi:10.1109/EDOC.2019.00021.
13. Korff, D.; Georges, M. The Data Protection Officer Handbook. SSRN Scholarly Paper ID 3428957, Social Science Research Network, Rochester, NY, 2019.
14. Pandit, H.J.; Polleres, A.; Bos, B.; Brennan, R.; Bruegger, B.; Ekaputra, F.J.; Fernández, J.D.; Hamed, R.G.; Lizar, M.; Schlehahn, E.; et al. Creating A Vocabulary for Data Privacy. In Proceedings of the The 18th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE2019); , 2019; p. 17. doi:10.1007/978-3-030-33246-4_44.
15. Data Catalog Vocabulary (DCAT) - Version 2.
16. DCAT-AP - Data.Gov.Ie. https://op.europa.eu/en/web/eu-vocabularies/dcat-ap.
17. Strategy for Data | Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/st data.
18. Measuring Privacy Operations 2019 Cookies, Local vs. Global Compliance, DSARs and More - IAPP and TrustArc. https://iapp.org/media/pdf/resource_center/measuring_privacy_operations_2019.
19. IAPP Tech Vendor Report 2020. https://iapp.org/media/pdf/resource_center/2020TechVendorReport.pd
20. The Value of Investing in Well-Constructed Records of Processing Activities - IAPP.
21. Sparapani, T.; Sherman, J. Privacy Tech's Third Generation A Review of the Emerging Privacy Tech Sector, 2021.
22. Khatri, V.; Brown, C.V. Designing Data Governance. *Communications of the ACM* **2010**, *53*, 148–152. doi:10.1145/1629175.1629210.
23. Rozehnal, P.; Novák, V. The Core Of Enterprise Architecture As A Management Tool: Gdpr Implementation Case Study. In Proceedings of the 26th Interdisciplinary Information Management Talks, 2020.
24. Burmeister, F.; Drews, P.; Schirmer, I. A Privacy-driven Enterprise Architecture Meta-Model for Supporting Compliance with the General Data Protection Regulation. In Proceedings of the Hawaii International Conference on System Sciences 2019 (HICSS-52), 2019.
25. Pandit, H.J.; Fatema, K.; O'Sullivan, D.; Lewis, D. GDPRtEXT - GDPR as a Linked Data Resource. In Proceedings of the The Semantic Web; Gangemi, A.; Navigli, R.; Vidal, M.E.; Hitzler, P.; Troncy, R.; Hollink, L.; Tordai, A.; Alam, M., Eds.; Springer International Publishing: Cham, 2018; Vol. 10843, pp. 481–495. doi:10.1007/978-3-319-93417-4_31.
26. Business Process Re-engineering and Functional Toolkit for GDPR Compliance (BPR4GDPR H2020 Project). https://www.bpr4gdpr.eu/.
27. Pandit, H.J.; Lewis, D. Modelling Provenance for GDPR Compliance Using Linked Open Data Vocabularies. In Proceedings of the Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn), 2017.
28. Pandit, H.J.; Debruyne, C.; O'Sullivan, D.; Lewis, D. GConsent - A Consent Ontology Based on the GDPR. In Proceedings of the The Semantic Web; Hitzler, P.; Fernández, M.; Janowicz, K.; Zaveri, A.; Gray, A.J.; Lopez, V.; Haller, A.; Hammar, K., Eds.; Springer International Publishing: Cham, 2019; Lecture Notes in Computer Science, pp. 270–282. doi:10.1007/978-3-030-21348-0_18.
29. Bonatti, P.A.; Kirrane, S.; Petrova, I.M.; Sauro, L. Machine Understandable Policies and GDPR Compliance Checking. *KI - Künstliche Intelligenz* **2020**, *34*, 303–315. doi:10.1007/s13218-020-00677-4.
30. Palmirani, M.; Martoni, M.; Rossi, A.; Bartolini, C.; Robaldo, L. PrOnto: Privacy Ontology for Legal Compliance. In Proceedings of the Proceedings of the 18th European Conference on Digital Government ECDG 2018, 2018, p. 10.
31. Lioudakis, G.V.; Koukovini, M.N.; Papagiannakopoulou, E.I.; Dellas, N.; Kalaboukas, K.; de Carvalho, R.M.; Hassani, M.; Bracciale, L.; Bianchi, G.; Juan-Verdejo, A.; et al. Facilitating GDPR Compliance: The H2020 BPR4GDPR Approach. In Proceedings of the Digital Transformation

for a Sustainable Society in the 21st Century; Pappas, I.O.; Mikalef, P.; Dwivedi, Y.K.; Jaccheri, L.; Krogstie, J.; Mäntymäki, M., Eds.; Springer International Publishing: Cham, 2020; IFIP Advances in Information and Communication Technology, pp. 72–78. doi:10.1007/978-3-030-39634-3_7.
32. Scharffe, F.; Fensel, D. Correspondence Patterns for Ontology Alignment. In Proceedings of the International Conference on Knowledge Engineering and Knowledge Management. Springer, 2008, pp. 83–92.
33. Brous, P.; Janssen, M.; Krans, R. Data Governance as Success Factor for Data Science. In Proceedings of the Responsible Design, Implementation and Use of Information and Communication Technology; Hattingh, M.; Matthee, M.; Smuts, H.; Pappas, I.; Dwivedi, Y.K.; Mäntymäki, M., Eds.; Springer International Publishing: Cham, 2020; pp. 431–442. doi:10.1007/978-3-030-44999-5_36.
34. Pandit, H.J.; O'Sullivan, D.; Lewis, D. Test-Driven Approach Towards GDPR Compliance. In Proceedings of the 15th International Conference on Semantic Systems (SEMANTiCS2019); , 2019. doi:10.1007/978-3-030-33220-4_2.
35. Coordinated Plan on Artificial Intelligence 2021 Review | Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review.
36. Ryan, P.; Crane, M.; Brennan, R. Design Challenges for GDPR RegTech:. In Proceedings of the Proceedings of the 22nd International Conference on Enterprise Information Systems; SCITEPRESS - Science and Technology Publications: Prague, Czech Republic, 2020; pp. 787–795. doi:10.5220/0009464507870795.
37. Arner, D.W.; Barberis, J.N.; Buckley, R.P. The Evolution of Fintech: A New Post-Crisis Paradigm? *SSRN Electronic Journal* **2015**. doi:10.2139/ssrn.2676553.
38. Privacy Impact Assessment (PIA) | CNIL. https://www.cnil.fr/en/privacy-impact-assessment-pia.

Appendix A CSM-ROPA: Analysis of ROPA requirements in GDPR and DPA templates

The following table summarises the comparison of information fields in GDPR and across DPA templates. The column 'GDPR' specifies the relevant clause, and 'Art.30' indicates whether the field is mandatory in a ROPA as per GDPR Art.30. The DPAs are denoted using the country's ISO Alpha-2 codes.

Table A1. Analysis of ROPA requirements in GDPR and DPA templates

GDPR	Field	A.30	BE	GR	GB	PL	CY	FR	PT	DE	DK	LU	FI	CZ	HR	IT	LT	SI	SK
5	Personal Data Location	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
5.1	Data Sources	×	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×
6.1	Legal basis	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6.1	Record of consent	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
9.1	Special Personal Data Category	×	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×
9.1	Vulnerable Data Subject Category	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
22.1	Automated decision making, profiling	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
26.1	Joint Controller agreement	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
28	Data Processors	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
28.3	Data Processing Contract	×	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
28.3	Data processor contract	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
30.1	Processing Status	×	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Tech/Org measures implementation	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Security measures	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Technologies used	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
33.5	Data Breach	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
35	Risk assessment and mitigation	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
35	DPIA Results	×	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×
35	Relevant DPIA	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

36.1	Impact Assessment, Prior Consultation	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
37.6	External DPO organisation	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
×	Business Process	×	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
×	Owner of Process	×	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×
×	Type of Processing	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
13, 14, 15	Data Subject Rights	×	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×
28, 30.1(c)	Third Party Data Transfer	×	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×
30.1(a)	Data Protection Officer Contact	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(a)	Representative	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(a)	Representative Contact	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(a)	Joint Controller Name	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(a)	Joint Controller contact	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(b)	Purposes of processing	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(b)	Main/Auxiliary Processing	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
30.1(c)	Personal Data Categories	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(c)	Data Subject Categories	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(d)	Recipient Categories	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(e)	Third Countries in Data Transfer	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(e)	Appropriate Safeguards	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30.1(f)	Retention/Deletion Periods	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
30.1(g)	Tech/Org measures	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30(1)(a)	Data Controller Contact	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30(1)(a)	Data Protection Officer	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
30(1)(a)	Data Protection Officer Contact	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
44-47	Nature of Transfer	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6.1(f)	Legitimate interests	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
6.1(f)	Legitimate interests assessment	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
6, 14, 30.1(b)	Data Combination	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	Nos. Fields	16	32	31	29	25	23	23	23	22	21	21	19	18	18	18	18	18	17

Appendix B CSM-ROPA: Mapping with DPV Concepts

The following table summarises the mapping between CSM-ROPA fields and DPV concepts. The column 'GDPR' specifies relevant clause in GDPR, 'DPV' specifies relevant concepts within DPV for expressing field information, 'Map.' refers to mapping outcome: *E* indicating Exact mapping i.e. the concept existed in DPV and could be used as is, *Pt* indicating Partial mapping i.e. the concept did not exist exactly, but another concept was similar in context, and *S* for indicating the concept did not exist and has been proposed for inclusion. The columns 'DC' and 'DP' represent the necessity of field for Data Controllers and Data Processors respectively, where: *M* indicates Mandatory i.e. a minimum requirement for ROPA as set out in article 30 , or as required for DPCAT functionality; *C* indicates Conditional i.e. a minimum requirement for ROPA as set under article 30 (if applicable); *R* indicates Recommended i.e. a non-legal requirement for ROPA that assists the organisation in meeting the accountability principle, recommended by DPA guidelines; and *O* indicates Optional i.e. a term found on a ROPA template that has no legal requirement for inclusion, nor any direct/ supplementary role in demonstrating accountability.

Table A2. Mapping of CSM-ROPA fields with DPV Concepts

GDPR	Field	DPV	Map.	DC	DP

5	Location of personal data	dpv:StorageLocation	E	R	R
5.1	Data Sources	dpv:DataSource	E	R	O
6.1	Legal basis	dpv:LegalBasis	E	M	O
6.1	Link to record of consent	dpv:Consent	E	R	R
9.1	Special Personal Data	dpv:SpecialCategoryPersonalData	E	R	O
9.1	Vulnerable Data Subjects	dpv:VulnerableDataSubject	E	R	O
22.1	Automated decision-making or profiling	dpv:AutomatedDecisionMaking	E	R	R
26.1	Joint Controller agreement	dpv:JointDataControllersAgreement	E	R	N/A
28	Data Processors	dpv:DataProcessor	E	R	M
28.3	Data Processing Contract	dpv:DataProcessingAgreement	E	R	R
28.3	Data processor contract	dpv:ControllerProcessorAgreement	E	R	R
30.1	Status of processing	dpv:Status	S	M	M
32	Tech/Org measures implementation	dpv:Technology	E	R	R
32	Security measures	dpv:TechnicalOrganisationalMeasure	E	R	R
32	Technologies used	dpv:Technology	E	R	R
33.5	Data Breach	dpcat:DataBreachRecord	S	R	R
35	Risk management	dpv:RiskMitigationMeasure	E	R	O
35	DPIA Results	dpv:DPIA	E	R	O
35	Relevant DPIA	dpv:DPIA	E	R	R
36.1	Impact Assessments	dpv:ImpactAssessment	E	R	R
36.1	Prior Consulatations	dpv:Consultation	E	R	R
37.6	External DPO organisation	dpv:DataProtectionOfficer	E	R	R
_	Name of Business Process	dpv:PersonalDataHandling	Pt	O	O
_	Owner of Process	dct:contactPoint	E	M	M
_	Type of Processing	dpv:Processing	E	O	O
13, 14, 15	Data Subject Rights	dpv:DataSubjectRight	E	R	O
28, 30.1(c)	Data Categories Transfer to Third Parties	dpv:Transfer, dpv:PersonalData	E	R	R
30.1(a)	DPO contact	dpv:hasName, dpv:hasContact	E	MC	MC
30.1(a)	Representative	dpv:Representative	E	MC	N/A
30.1(a)	Representative contact	dpv:hasName, dpv:hasContact	E	MC	N/A
30.1(a)	Name of joint controller	dpv:JointDataController	E	MC	N/A
30.1(a)	Contact details of joint controller	dpv:hasName, dpv:hasContact	E	MC	N/A
30.1(b)	Purposes of processing	dpv: Purpose	E	M	O
30.1(b)	Main/Auxilary Processing	dpv:Importance (Primary, Secondary)	E	O	O
30.1(c)	Personal Data Categories	dpv:PersonalDataCategory	E	M	M
30.1(c)	Categories of data subjects	dpv:DataSubject	E	M	M
30.1(d)	Categories of Recipients	dpv:Recipient	E	MC	MC
30.1(e)	Third Countries Data Transfer	dpv:ThirdCountry	E	MC	MC
30.1(e)	Appropriate Safeguards	dpv:Safeguard	E	MC	MC
30.1(f)	Retention/Deletion Periods	dpv:StorageDuration,	E	M	O
30.1(g)	Technical and organisational measures	dpv:TechnicalOrganisationalMeasure	E	M	M
30(1)(a)	Data Controller contact	dpv:hasName, dpv:hasContact	E	M	M
30(1)(a)	Data Protection Officer	dpv:DataProtectionOfficer	E	MC	MC
44-47	Nature of Transfer	dpv:DataTransferLegalBasis	E	MC	MC
6.1(f)	Legitimate interests	dpv:LegitimateInterest	E	R	R
6.1(f)	Legitimate interests assessment	dpv:LegitimateInterestAssessment	E	R	R
6, 14, 30.1(b)	Data Combination	dpv:Combine	E	R	O

Appendix C DPCat Specification

The following tables summarise the *ROPA*, *ROPACatalog*, and *ROPARecord* fields in the DPCat specification. The 'Card.' columns refer to the cardinality of the field, and 'Nec.' columns refer to necessity requirements for the fields, where: *M* indicates Mandatory; *C* indicates Conditional i.e. if applicable; *R* indicates Recommended; and *O* indicates Optional.

Table A3. DPCat *ROPA* and *ROPACatalog* fields

Title	Relation	Domain	Range	Card.	Nec.
Datasets in Catalog	dcat:dataset	dpcat:ROPA(Catalog)	dpcat:ROPARecord	0...n	M
Description	dct:description	dpcat:ROPA(Catalog)	rdfs:Literal	1...n	M
Issued	dct:issued	dpcat:ROPA(Catalog)	rdfs:Literal (XSD date/time)	0...1	R
Publisher	dct:publisher	dpcat:ROPA(Catalog)	foaf:Agent	1..1	M
Title	dct:title	dpcat:ROPA(Catalog)	rdfs:Literal	1...n	M
Contact Point	dcat:contactPoint	dpcat:ROPA(Catalog)	vcard:Kind	0..n	R
Temporal coverage	dct:temporal	dpcat:ROPA(Catalog)	dct:PeriodOfTime	0..n	O
Data Controller	dpv:hasDataController	dpcat:ROPA(Catalog)	dpv:DataController	1...1	M
DPO for Catalog	dpv:hasDataProtectionOfficer	dpcat:ROPA(Catalog)	dpv:Data ProtectionOfficer	0...1	MC
Representative	dpv:hasRepresentative	dpcat:ROPA(Catalog)	dpv:Representative	0...1	MC
Responsible Entity	dpcat:responsible Entity	dpcat:ROPA(Catalog)	dpv:Entity	0...n	O
Catalogs	dcat:catalog	dcat:ROPACatalog	dpv:ROPA	0..n	M

Table A4. DPCat *ROPARecord* fields

Title	Relation	Domain	Range	Card.	Nec.
Contract Point	dcat:contactPoint	dpcat:ROPARecord	vcard:Kind	0...n	R
Description	dct:description	dpcat:ROPARecord	rdfs:Literal	1..n	M
Identifier	dct:identifier	dpcat:ROPARecord	rdfs:Literal	0...n	O
Date Issued	dct:issued	dpcat:ROPARecord	rdfs:Literal (datetime)	0..1	O
Publisher	dct:publisher	dpcat:ROPARecord	foaf:Agent	0..1	R
Temporal coverage	dct:temporal	dpcat:ROPARecord	dct:PeriodOfTime	0..n	R
Title	dct:title	dpcat:ROPARecord	rdfs:Literal	1..n	M
Joint Controller	dpv: hasJointDataControllers	dpcat:ROPARecord	dpv:LegalEntity	0..n	MC
Business Process	dpv:hasPersonalDataHandling	dpcat:ROPARecord	dpv:Personal DataHandling	0..1	R
Process Owner	dcat:contactPoint	dpcat:ROPARecord	vcard:Kind	0..n	R
Purposes	dpv:hasPurpose	dpcat:ROPARecord	dpv:Purpose	1..n	M
Legal Basis	dpv:hasLegalBasis	dpcat:ROPARecord	dpv:LegalBasis	1..n	M
Type of Processing	dpv:hasProcessing	dpcat:ROPARecord	dpv:Processing	1..n	R
Personal Data	dpv:hasPersonalData	dpcat:ROPARecord	dpv:PersonalData	1...n	M
Special Personal Data Categories	rdfs:subClassOf	dpv:SpecialCategory Personaldata	dpv:PersonalData	1...n	R
Data Subjects	dpv:hasDataSubject	dpcat:ROPARecord	dpv:DataSubject	1..n	M
Vulnerable Data Subjects	rdfs:subClassOf	dpv:Vulnerable DataSubject	dpv:DataSubject	0..n	R
Data Retention / Deletion Periods	dpv:hasStorage	dpcat:ROPARecord	dpv:StorageDuration	1...n	M
Data Combination	rdfs:subClassOf	dpv:Combine	dpv:Processing	0..n	R
Source of Data	dpv:hasDataSource	dpcat:ROPARecord	dpv:DataSource	1..n	R
Processor	dpv:hasDataProcessor	dpcat:ROPARecord	dpv:LegalEntity	0..n	M
Data Processing Contract	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv: DataProcessingAgreement	0..n	R
Recipients	dpv:hasRecipient	dpcat:ROPARecord	dpv:LegalEntity	1...n	MC
Third countries for Transfers	dpv:hasThirdCountry	dpv:Transfer	dpv:ThirdCountry	0..n	MC
Nature of Transfer	dpv:hasLegalBasis	dpv:Transfer	dpv:LegalBasis	0..n	MC
Safeguards	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:Safeguard	0..n	MC
Risk management	dpv:hasRisk, dpv:isMitigatedByMeasure	dpcat:ROPARecord	dpv:Risk, dpv:RiskMitigationMeasure	0..n	R
Technical / Organisational measures	dpv:hasTechnicalOrganisational Measure	dpcat:ROPARecord	dpv:TechnicalOrganistional Measure	1...n	M
DPIA	dpv:hasOrganisational Measure	dpcat:ROPARecord	dpv:DPIA	0..n	R
Data Subject Rights	dpv:hasRight	dpcat:ROPARecord	dpv:DataSubjectRight	1..n	R
Legitimate interests	dpv:hasLegalBasis	dpcat:ROPARecord	dpv:LegitimateInterest	0..n	R
Legitimate Interests Assessment	dpv:hasOrganisational Measure	dpv: LegitimateInterest	dpv:LegitimateInterest Assessment	0..n	R
Automated decision-making	dpv:hasContext	dpv:Processing	dpv:AutomatedDecision Making	0..n	R
Profiling	rdfs:subClassOf	dpv:Profiling	dpv:Processing	0..n	R
Record of Consent	dpv:hasLegalBasis	dpcat:ROPARecord	dpv:Consent	0..n	R
Location of Personal Data	dpv:hasStorage	dpcat:ROPARecord	dpv:StorageLocation	1..n	R
Status of Processing	dpv:hasContext	dpcat:ROPARecord	dpv:Status	1..n	R
Relevant Personal Data Breach	dpcat:associatedWithDataBreach	dpcat:ROPARecord	dpcat:DataBreach	0..n	R
Impact Assessment	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:ImpactAssessment	0..n	R
Prior Consulatation	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:Consultation	0..n	R
Main/Auxilary Processing	dpv:hasContext	dpcat:ROPARecord	dpv:Importance	1..n	R
Joint Controller Agreement	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:JointDataControllers Agreement	0..n	R
Data Processor Contract	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:ControllerProcessor Agreement	0..n	R
Information System for Tech/Org measure	dpv: isImpementedUsingTechnology	dpv: TechnicalOrganisationalMeasure	dpv:Technology	1..n	R
Security Measures	dpv:hasTechnicalOrganisational Measure	dpcat:ROPARecord	dpv:TechnicalOrganisationalMeasure	1..n	R
Relevant DPIA	dpv:hasOrganisationalMeasure	dpcat:ROPARecord	dpv:DPIA	0..n	R
System or software name	dpv:isImpementedUsing Technology	dpcat:ROPARecord	dpv:Technology	1..n	R