ERoPA: A Machine-Readable Approach to the Record of Processing Activities (RoPA) for GDPR Compliance

A thesis submitted to the
Dublin City University
for the degree of
Doctor of Philosophy

Paul Ryan B. Comm, MSc.

School of Computing
Dublin City University
Ireland
Supervisors: Asst. Prof. Harshvardhan J Pandit, Prof. Martin Crane, and Asst. Prof. Rob Brennan (University College Dublin)
January 2026

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. I hereby certify that no Generative Artificial Intelligence (Gen AI) tools have been used in the creation of the thesis.

Paul Ryan (Candidate)
ID No: 19214306
Date: Dec 4th, 2025

Acknowledgements

I would like to thank my supervisor, Dr Rob Brennan, for his guidance, encouragement, and support over the years. I thank my additional supervisors, Dr. Harshvardhan Pandit and Dr. Martin Crane, for their valuable suggestions and great patience. A warm thank you to all the Data Privacy Vocabulary Community Group members and the ADAPT Transparent Digital Governance research strand, from whom I have learned much. Thank you to Dublin City University, the SFI Adapt Research Centre, and Uniphar PLC for funding this research. My greatest appreciation goes to all my wonderful friends who have shown much love and support. I am most grateful for my family's unconditional support and unwavering belief, to whom this thesis is dedicated.

Table of Contents

Declaration II
Acknowledgements III
Table of Contents IV
Table of Figures VII
Table of Tables VIII
Table of Listings IX
Abbreviations X
Glossary XI
Abstract XIII

1 INTRODUCTION 1
1.1 Motivation 1
1.2 Research Question 4
1.3 Research Sub Questions 4
1.4 Methodology 5
1.5 Technical Approach 7
1.6 Evaluation Approach 7
1.7 Contributions 9
1.8 Thesis Overview 11

2 STATE OF THE ART 14
2.1 Chapter Overview 14
2.2 An Overview of the GDPR Accountability Principle 14
2.3 The Record of Processing Activities (RoPA) under the GDPR 16
2.3.1 The RoPA as an Integral Component of GDPR Accountability 16
2.3.2 Information Present in a RoPA 17
2.3.3 Stakeholder Roles regarding RoPA. 20
2.3.4 RoPA and the Role of the Data Protection Officer (DPO) 21
2.4 Current Organisational Approaches to GDPR Compliance 22
2.4.1 Manual and Informal Approaches to GDPR Compliance 22
2.4.2 Enterprise Software Solutions 23
2.4.3 Maturity Models 27
2.4.4 Accountability Frameworks 28
2.4.5 GDPR Certification 29
2.4.6 Overview of Existing Organisational Approaches to GDPR Compliance 30
2.5 Technological Approaches to GDPR Compliance 32
2.5.1 The Success Factors of RegTech 32
2.5.2 RegTech for Supporting GDPR Compliance 35
2.5.3 Standard Initiatives to GDPR Compliance 39
2.5.4 Semantic Web Approaches to GDPR Compliance 41
2.5.5 Overview of Technical Approaches to RoPA Maintenance 43
2.6 Conclusion 45

3 RESEARCH METHODOLOGY AND BACKGROUND 47
3.1 Research Approach 47
3.2 Research Design 49
3.2.1 Design for the Problem Formulation Stage 49
3.2.2 Design for Building, Intervention, and Evaluation (BIE) Stages 49
3.2.3 Design for the Formalisation of the Learning Stage. 58
3.3 Data Collection and Analysis Methods 59
3.4 Ethical Considerations: 60
3.5 Limitations 60

3.6 Background Information 61
3.7 Chapter Summary 64

4 REQUIREMENTS FOR THE EROPA APPROACH 66
4.1 Chapter Overview 66
4.1.1 ADR Roles for this Chapter 66
4.2 Requirements Gathering Process for the ERoPA Approach 67
4.3 Requirements Elicitation Techniques 68
4.4 Gathering the Requirements for Machine-Readable RoPA 69
4.5 Analysis of the Authoritative Sources for Requirements for ERoPA 69
4.5.1 Legal Texts 69
4.5.2 Regulator RoPA Templates 71
4.5.3 Regulator Guidance on RoPA 77
4.6 Analysis of Commercial Practice to Identify Requirements for ERoPA 79
4.6.1 Current Commercial Tools and Initiatives 79
4.6.2 Survey of Data Protection Professionals 80
4.7 Specification of the Requirements for ERoPA Approach 87
4.8 Traceability Matrix of ERoPA Requirements 89
4.9 Summary of ERoPA Requirements Analysis 90

5 THE COMMON SEMANTIC MODEL OF ROPA (CSM-ROPA) 92
5.1 Chapter Overview 92
5.1.1 ADR Roles 92
5.2 CSM-RoPA Design 93
5.2.1 Ontology Engineering Methodology 93
5.2.2 Requirement Specification for CSM-RoPA 95
5.2.3 Knowledge Acquisition 98
5.2.4 Ontology Reuse and Reengineering 101
5.2.5 Ontology Design 107
5.3 Ontology Implementation 114
5.4 Evaluation 114
5.4.1 Answering the Competency Questions 115
5.4.2 Following Ontology Engineering Best Practices 116
5.4.3 Use Case – Representing a RoPA 117
5.4.4 Peer Review in W3C DPVCG 118
5.4.5 Peer-Reviewed Publications and Industry Uptake 119
5.5 Learning from the Development and Evaluation of the Semantic Model 119
5.6 Conclusions 120

6 INTEROPERABLE ROPA SPECIFICATION 121
6.1 Chapter Overview 121
6.1.1 ADR Roles 122
6.2 Use Cases 122
Use Case U1: Data Controller 124
Use Case U2 Data Controller with Internal Organisational Units 124
Use Case U3: Data Controller with Data Processors 125
Use Case U4: Data Controller in a Joint Controller Relationship 126
Use Case U5: DPO Overseeing Multiple Data Controllers 127
6.3 DPCat Design 128
6.3.1 Interoperability Specification for RoPA Information 129
6.3.2 Knowledge Acquisition 131
6.3.3 Ontology Re-use and Engineering 132
6.4 DPCat Implementation 133
6.5 DPCat Evaluation 137
6.5.1 Case Study Technical Approach 138

6.5.2 Case Study Set-up ... *138*
6.5.3 Case Study Results ... *144*
6.5.4 Analysis ... *145*
6.6 Learnings from the Development and Evaluation of DPCat ... 148
6.7 Conclusions ... 149
6.7.1 Peer-Reviewed Publications ... *149*
7 THE EROPA APPROACH AND UPSILON CASE STUDY ... 151
7.1 Chapter Overview ... 151
7.1.1 ADR Roles ... *152*
7.2 The ERoPA Approach ... 153
7.2.1 Tools and Methods: ... *154*
7.2.2 The Role of Stakeholders in the ERoPA Approach ... *155*
7.2.3 System Capabilities ... *156*
7.3 Upsilon Deployment Case Study ... 158
7.4 Case Study Design ... 160
7.4.1 Design for Technical Activities ... *160*
7.4.2 Design for Observational Feedback on Deployment ... *164*
7.4.3 Design for Expert Feedback ... *164*
7.4.4 Design for ERoPA Accountability Verification ... *166*
7.4.5 Presentation of Findings ... *168*
7.5 Data Collection ... 169
7.5.1 Deployment Process ... *169*
7.5.2 Observations made through the deployment process ... *174*
7.5.3 Semi-structured Interview Data ... *176*
7.5.4 ERoPA Accountability Verification Data ... *177*
7.6 Data Analysis ... 181
7.6.1. Analysis of Direct Observations of Deployment ... *181*
7.6.2 Analysis of Expert Feedback from Semi-structured Interviews ... *182*
7.6.3 Analysis of ICO Accountability Framework Verification using ERoPA ... *183*
7.6.4 Case Study Synthesis ... *185*
7.7 Case Study Findings ... 186
7.8 ERoPA Deployment Guidelines ... 187
7.8.1 Methodology ... *188*
7.8.2 Data Collection ... *189*
7.8.3 Data Analysis: ... *190*
7.8.4 Approach for Implementing ERoPA Zachman Framework ... *191*
7.9 Chapter Conclusion ... 195
8 CONCLUSIONS ... 196
8.1 Chapter Overview ... 196
8.2 Summary of the Key Findings ... 196
8.3 Responding to the Research Question ... 197
8.4 Reflection on the ADR Process ... 200
8.5 Contributions ... 202
8.6 Limitations ... 203
8.7 Future Work ... 203
REFERENCES ... 205
APPENDICES ... 216
Appendix A Survey of Data Protection Professionals ... *216*
Appendix B Mapping of GDPR RoPA Concepts to DPV terms (2024) ... *221*
Appendix C CSM-RoPA 0.3 Classes and Relationships 2024 ... *223*
Appendix D Serialised RDF representation of Data Protection Officer Class. ... *231*
Appendix E Serialised RDF representation of Object Property. ... *232*

Appendix F How Competence Questions are met using terms from CSM-RoPA..... 233
Appendix G SPARQL Query to Represent RoPA with CSM-RoPA..... 235
Appendix H Sample Extracts taken from EDPS RoPA..... 236
Appendix I Mapping of ICO concepts to DPV terms..... 237
Appendix J Sample RDF Listing Generated from OntoRefine..... 243
Appendix K DPCat RDF file prepared by Third Party Processor..... 244
Appendix L Semi-Structured Interviews..... 245

Table of Figures

Figure 1 Action Design Research Methodology Approach Overview 6
Figure 2 Stakeholder Contributions of GDPR Accountability Documents to RoPA. 17
Figure 3 An Example of a RoPA Provided by the Data Protection Commission [11] 19
Figure 4 Primary Tools for Data Inventory and Mapping (IAPP) [12] 22
Figure 5 Primary Solution Used to Manage Privacy Program in 2020-21 (IAPP) [19]..... 23
Figure 6 Growth of Privacy Technology Marketplace [62] 24
Figure 7 The Accountability Wheel/Universal Elements of Accountability from CIPL [42] 28
Figure 8 Overview of Number of Regulator Templates vs. No of Fields Instances Found..... 76
Figure 9 Survey Results - How challenged are organisations with Maintaining RoPA?..... 83
Figure 10 Stakeholder Relationship Gaining Most Benefit from ERoPA..... 84
Figure 11 Stakeholder Relationship Gaining the Second Most Benefit from ERoPA..... 85
Figure 12 Where will the DPO Use ERoPA? 86
Figure 13 First Build, Implement and Evaluate Stage of ERoPA Development..... 92
Figure 14 UML Representation of the ROPA Model based on Regulator Templates. 100
Figure 15 Classes and Subclasses to Represent Processing from DPV [6]. 108
Figure 16 A Graphical Representation of CSM-RoPA 0.3..... 113
Figure 17 Second Build, Implement and Evaluate Stage of ERoPA Development..... 121
Figure 18 Basic generation of legal requirement ROPA [18]. 124
Figure 19 Organisational units updating and maintaining RoPA [18]..... 125
Figure 20 A data controller with organisational units and data processors [18]..... 126
Figure 21 Data controller in a joint controller relationship [18]..... 126
Figure 22 A DPO Overseeing Multiple Data Controllers [18]..... 127
Figure 23 DPCat specification for interoperability of GDPR information. 135
Figure 24 DPCat Case Study Technical Approach..... 138
Figure 25 Action Design Research showing the Stages covered by this chapter..... 151
Figure 26 Overview of the Tools and Methods Used in the ERoPA Approach. 153
Figure 27 Extended Capability Model for Data Management in GDPR [18] 156
Figure 28 Technical Design for ERoPA Case Study..... 160
Figure 29 Visualisation of a RoPA Record Fragment an ERoPA Knowledge Graph..... 171
Figure 30 SPARQL Query Output Identifying Non-Compliant RoPA Record..... 173
Figure 31: Extent of DPCat coverage to represent the ICO Accountability Framework..... 178
Figure 32 Analysis of Themes from Semi-structured Interviews..... 182

Table of Tables

Table 1 Overview of Which Thesis Chapters Address Each Research Sub Question 5
Table 2 Definition and Allocation of Action Design Research Roles. 6
Table 3 Artefact Component Parts Produced as part of this Research 11
Table 4 Privacy Software Tools, Number of Vendors per Category [60], [63] 25
Table 5 A Comparison of the Capabilities for Existing GDPR Compliance Approaches. 31
Table 6 Features of RegTech Systems 34
Table 7 GDPR Compliance Automation - Analysis of the Primary Approach Taken 36
Table 8 Review of State-of-the-Art Approaches to RoPAs 44
Table 9 Overview of the ICO Accountability Framework 64
Table 10 ADR Roles for the Problem Formulation Stage 66
Table 11 Sources used to Gather the Requirements for ERoPA 69
Table 12 An Analysis of Concepts Found in GDPR Art. 30 Text 70
Table 13 Analysis of RoPA Requirements in GDPR and DPA Templates. 72
Table 14 Qualifications of Respondents to Survey. 82
Table 15 Requirements Traceability Matrix. 90
Table 16 ADR Role Assignment for BIE1 93
Table 17 CSM-RoPA Ontology Engineering Workflow Steps. 94
Table 18 Competence Questions that CSM-RoPA must meet based on GDPR Art.30. 96
Table 19 Terms Identified in Regulator Templates and Frequency of Occurrence. 97
Table 20 Supplementary Competence Questions gathered from Regulator Templates 98
Table 21 Data Sources for CSM-RoPA Knowledge Acquisition Stage. 98
Table 22 Example of GDPR Concept Matching Process with DPV Ontology. 102
Table 23 First Mapping GDPR concepts to DPV terms (2020). 102
Table 24 Second Mapping of GDPR concepts to DPV terms (2022). 104
Table 25 Summary of Outcomes of Mapping Cycles between GDPR Concepts and DPV 105
Table 26 Mapping of CSM-RoPA with DPV Concepts. 106
Table 27 Competence Questions from Ontology Engineering Requirements 107
Table 28 Overview of Evolution of CSM-RoPA Mapping Outcomes. 109
Table 29 Sample of CSM-RoPA 0.3 Classes and Relationships 2024 112
Table 30 Meeting the Competence Questions using terms from CSM-RoPA 115
Table 31 Sample Extract of Controller ROPA. 118
Table 32 Peer-Reviewed Publications concerning the Semantic Model of RoPA. 119
Table 33 ADR Role Assignment for BIE 2 122
Table 34 DPCat Specification Build Steps. 128
Table 35 DPCat Definition of Terms between DCAT-AP and DPCat 134
Table 36 DPCat ROPA and ROPACatalog fields. 136
Table 37 Sample of DPCat ROPARecord Fields. 137
Table 38 Cross-reference of Use cases and Relevant EDPS RoPA Records. 139
Table 39 Query Output of Article 30 Format RoPA for Regulator. 142
Table 40 Query Results for an Overview of Processing for DPO using DPCat. 143
Table 41 Summary of Outcomes from DPCat Case Study Scenarios. 144
Table 42 Extent that DPCat meets the Competence Questions. 146
Table 43 Publications and conference presentations concerning DPCat 149
Table 44 ADR Role Assignment for BIE 3 152
Table 45 Role Played in ERoPA by Stakeholders. 155
Table 46 ERoPA System capabilities and specific requirements for ERoPA 157
Table 47 Question Set Utilised in Semi-Structured Interviews. 165
Table 48 Description of Dataset Used for Case Study. 169
Table 49 Example of Ontorefine Mapping to ERoPA 171
Table 50 Sample of SPARQL Report used to Identify Conflict between RoPA Records. 173
Table 51 Observations Gathered through the Upsilon ERoPA Case Study. 174

Table 52 ICO terms to ERoPA Mapping. 178
Table 53 Extent to which the Upsilon ERoPA meets Regulator Expectations. 179
Table 54 Synthesis of Case Study Data Analyses 185
Table 55 ERoPA Findings Gathered in this Thesis. 189
Table 56 Guidelines for organisations deploying ERoPA. 193

Table of Listings

Listing 1 Representation of ROPA and ROPARecord for EDPS Document. 140
Listing 2 SPARQL Query to Generate Article 30 RoPA. 142
Listing 3 SPARQL Query for an Overview of Processing for DPO using DPCat. 143
Listing 4 SPARQL Query Output Identifying a Conflict on a RoPA record. 163
Listing 5 SPARQL Query to identify Records without a Legal Basis Entry 163

Abbreviations

ADR	Action Design Research
API	Application Programming Interface
BIE	Building, Intervention and Evaluation stage of ADR methodology
CSM-RoPA	Common Semantic Model of RoPA
DPC	Data Protection Commission
DPCat	Data Processing Catalogue specification for GDPR RoPA information exchange
DPA	Data Processing Agreement
DPIA	Data Privacy Impact Assessment
DPO	Data Protection Officer
DPV	Data Privacy Vocabulary
DPVCG	W3C Data Privacy Vocabularies and Controls Community Group
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
ERoPA	Electronic Record of Processing Activities
GDPR	General Data Protection Regulation
IAPP	International Association of Privacy Professionals
ICO	Information Commissioner's Office (United Kingdom)
IPEN	Internet Privacy Engineering Network
RoPA	Record of Processing Activities
W3C	World Wide Web Consortium

Glossary

Action Design Research is an information systems methodology that builds the artefact through interactions between researchers and practitioners within the organisational context [1].

Artefact: An object made by humans with the intention that it is used to address a practical problem [2].

Capability: An organisational capability is the ability of a business to perform a coordinated task, utilising organisational resources, to achieve a particular result [3], [4].

Concept: In the Semantic Web, a concept usually denotes an abstract or concrete idea or a classification of items that can be expressed and analysed in a manner that machines can interpret. Concepts assist in providing significance to data and enable computers to comprehend the connections between pieces of information [5].

Data Controller: This is an entity that is the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of processing personal data (see Art[1] 4.7).

Data Processor: This is an entity that processes personal data solely on behalf of the data controller. The data processor is usually a third-party external to an organisation (see Art. 4.8).

Data Protection Officer: This role independently ensures that an organisation applies laws protecting individuals' personal data. The designation, position, and tasks of a DPO within an organisation are defined in the European Union General Data Protection Regulation (see Art. 37-39).

Data Privacy Vocabulary: (DPV) is an ontology of concepts for the interoperable representation and exchange of information about processing of personal data and the use of technologies. The DPV is serialised using semantic-web standards to represent privacy and data protection concepts primarily derived from GDPR [6]. The DPV is helpful as a machine-readable representation of personal data processing and can be adopted in relevant use cases such as legal compliance documentation and evaluation, policy specification, consent representation and requests, taxonomy of legal terms, and annotation of text and data [6].

DPCat (Data Processing Catalogue): A specification for an interoperable and machine-readable catalogue of data processing activities within one or more organisations based on the requirements of the General Data Protection Regulation (GDPR) and EU DPA guidelines. It extends the Data Catalog Vocabulary (DCAT) - Version 2 and the DCAT Application profile for data portals in Europe

[1] Art. refers to the relevant GDPR Article section

(DCAT-AP) standards. It reuses the Data Privacy Vocabulary (DPV) Specification to enable data governance of RoPA and related information across various use cases.

Entity: In the Semantic Web, an entity typically denotes a distinct, recognisable item in the world that can be represented through data. It serves as an instance of a concept or category [7].

Machine-readable RoPA: records processing activities created and maintained with computer-readable data or data presented in a format that a computer can process.

Practitioner: A practitioner is someone who is actively involved in the domain being studied. They actively engage in designing, developing, and evaluating solutions related to their practice. The practitioner is an active participant in the research process and brings practical knowledge, experience, and understanding of the problem, while also learning and adapting their practice based on research findings [1].

Record of Processing Activities It is a legal requirement in Article 30 of the General Data Protection Regulation (GDPR). It requires data controllers to maintain a record of processing activities (RoPA), which is their responsibility. Article 30 of GDPR prescribes the information the records must contain and states that controllers and processors must be able to provide such records to the Data Protection Commission (DPC) on request. The Records of Processing Activities (RoPA), as a measure to demonstrate compliance, can support Data Controllers demonstrate and implement the principle of accountability as set out in Article 5(2) GDPR. A well-drafted RoPA will explain to the DPC that a Data Controller is aware of and has considered the purpose of all processing activities within the organisation.

Semantic Web is a framework designed to help machines process and interpret data on the Internet by organising it in a machine-readable and interoperable format. It utilises technologies such as RDF (Resource Description Framework), OWL (Web Ontology Language), and SPARQL (SPARQL Protocol and RDF Query Language) to represent relationships between data, define specific ontologies for different domains, and query structured information. This approach enhances intelligent data integration, reasoning, and automation by offering a standardised format and protocols for machines to process web content and metadata, moving beyond simple keywords or natural language text. Ultimately, it transforms data into a web of connected, meaningful information[5], [8] .

User: A user is typically seen as an individual, a group, or an organisation that will directly use or benefit from the outcome, output, or results of the research [1].

Abstract

ERoPA: A Machine-Readable Approach to the Record of Processing Activities (RoPA) for GDPR Compliance.

Paul Ryan

The General Data Protection Regulation (GDPR) mandates that organisations keep a Record of Processing Activities (RoPA) and ensure compliance. The RoPA should include details on processing personal data from internal departments with diverse IT systems and external data processors. Current practices rely on spreadsheets or proprietary systems, which lack machine readability and interoperability, creating obstacles to automation. Regulators report that organisations face challenges in maintaining an accurate and up-to-date RoPA.

This thesis defines an approach to supporting 'Electronic Records of Processing Activities' (ERoPA) to help organisations comply with the GDPR Accountability Principle. The "ERoPA Approach" facilitates the collection, representation, transfer, and review of information to support organisational GDPR compliance through the automation of RoPA processes based on stakeholder requirements.

Using the Action Design Research (ADR) methodology, fourteen stakeholder requirements for the ERoPA Approach were identified. The ERoPA Approach was developed iteratively through ADR to provide: (i) an ontology to support the representation of RoPAs based on a survey of RoPA templates published by GDPR regulators, (ii) an interoperable machine-readable approach for the collection and transfer of RoPA information, and (iii) queries to support typical compliance tasks and (iv) deployment guidelines for practical implementations based on a case study in a real organisation where observations were gathered, and the opinions of data protection experts were consulted.

The main contribution of this thesis is the ERoPA Approach, which enhances GDPR accountability by facilitating the collection, representation, transfer, and review of RoPA information exchanged among stakeholders in data processing chains. The ERoPA Approach enables sharing GDPR accountability information with regulators and certification bodies, significantly improving the visibility and efficiency of organisational accountability practices. Additionally, it provides tools to support GDPR compliance automation. A minor contribution of this research is the extension of the W3C Community Standard, Data Privacy Vocabulary (DPV), to represent RoPAs.

1 Introduction

1.1 Motivation

The need for organisations to ensure that they are compliant with the General Data Protection Regulation (GDPR) has never been more critical [9] as fines have reached over four billion euros across 2200 different organisations[2].

The GDPR requires organisations to show that their processing activities are legal, fair, and transparent and that the data is adequately protected, accurate, and processed only for as long as necessary (see GDPR Article 5, hereafter abbreviated as Art.5). The GDPR applies to the processing of personal data and requires that the controllers and processors consider the nature, scope, context, and purposes of the processing and the risks to the rights and freedoms of natural persons to implement appropriate technical and organisational measures to ensure and demonstrate compliance (see Art.32).

To achieve this, organisations must ensure that they have adequate documentation concerning their data processing activities so that they can be informed of any risks related to the processing (see Art.5). Many organisations may appoint a data protection officer (DPO) as required by the GDPR, to monitor, advise and inform them regarding the organisation's compliance (see Art.37). The DPO is also responsible for maintaining a Record of Processing Activities (RoPA) which is a mandatory GDPR compliance document (see Art.30) [10]. The RoPA provides the organisation's DPO and regulators with a comprehensive overview of all personal data processing carried out by the organisation. Completing a RoPA involves gathering specific information from stakeholders, such as departments, organisational units, and (external to the organisation) data processors, and recording this information in a document. A comprehensive and well-formed RoPA is thus a vital resource that assists the DPO in assessing the organisation's GDPR compliance [11].

In 2022, the Irish Data Protection Commission (DPC) inspected the RoPAs of 30 organisations across public and private enterprises [11]. The review identified several common failures in maintaining a RoPA. Among the shortcomings identified in the report were that organisations' RoPAs needed to be more accurate, more detailed, and up to date. The report states that the best practice for RoPA requires an electronic living and dynamic document, which requires regular engagement and is updated and reviewed frequently to reflect the organisation's current compliance status [11].

[2] https://www.enforcementtracker.com/ (May 2024)

The approach that many organisations take to RoPA maintenance is to manually gather an inventory of processing activities and record this information in stand-alone spreadsheets based on templates provided by regulators [12]. As modern organisations tend to have a complex structure of organisational units supported by multiple outsourced providers, gathering and representing up-to-date and accurate data for RoPA is challenging [11]. Many stakeholders conduct diverse personal data processing activities where the GDPR compliance information associated with the processing activity exists in many forms, such as data processing agreements, privacy notices and enterprise architecture models.

The organisation's challenge is that current RoPA processes are often manual, resulting in out-of-date and inaccurate RoPA records. There is a need for an automated, interoperable electronic RoPA that uses standardised i.e. consistent representations for information from multiple sources to support the DPO's tasks in creating, validating, maintaining, and using this information to implement GDPR compliance and accountability within the organisation.

To address this need for an electronic RoPA, privacy software vendors offer solutions for RoPA management, often as part of a more extensive suite of GDPR compliance tools [13], [14]. This move towards electronic RoPAs follows the increasing trend of organisations adopting regulatory technology (often called *RegTech*)[3] [15] to assist with legal compliance and requirements. Vendor-supplied privacy software solutions provide online repositories for managing the RoPA within their proprietary software systems. These software systems often provide upload tools or APIs for onboarding a new client's RoPA or outputting a RoPA generated for presentation to a regulator. While the movement of organisations to use electronic RoPAs is a positive step towards automating GDPR compliance, the need to standardise RoPA accountability information presents interoperability challenges between stakeholders. This means organisations using such proprietary RoPA solutions can find connecting the RoPA data to their other systems challenging without explicit (and potentially expensive) vendor support. This results in a loss of opportunity and potential innovation and creates a lock-in effect where the organisation cannot use any other tool or service for GDPR interoperability. This lack of interoperability presents a challenge for all new vendors providing such privacy software services.

To address this challenge, this research builds on the existing successes of RegTech solutions in non-GDPR domains [15], [16]. The use of standardised, interoperable metadata has proven to be a key driver for compliance using Regtech [15], and this approach extends readily to create RegTech solutions specific to the GDPR [17]. Such a standardised approach to RoPA

[3] RegTech is information technology (IT) that (a) helps firms manage regulatory requirements and compliance imperatives by identifying the impacts of regulatory provisions on business models, products and services, functional activities, policies, operational procedures and controls; (b) enables compliant business systems and data; (c) helps control and manage regulatory, financial and non-financial risks; and (d) performs regulatory compliance reporting [15]

information would provide a practical benefit to stakeholders. For organisations, the benefits are consistent quality, trust, improved regulatory control, error reduction, lower cost of compliance, and confidence that processes are conducted consistently. For DPOs, it would provide them with up-to-date and accurate information regarding the processing activities.

A standardised machine-readable approach to GDPR compliance, where data can easily be processed by a computer without human intervention while ensuring no semantic meaning is lost, requires consistent representation and interpretation, which is vital in legal compliance processes[18]. Semantic Web standards are useful here as they provide standardised, interoperable formats and protocols to process content and metadata beyond simple keywords or natural language text [5] to transform data into a 'graph' of connected web of meaningful knowledge (i.e. a knowledge graph). The approach of utilising the Semantic Web to automate the gathering and updating of heterogeneous information offers organisations the potential for an accurate, comprehensive, and up-to-date electronic RoPA.

A semantic model of the RoPA has the potential to build trust and confidence in the data processing chain, as it involves multiple actors and heterogeneous sources of information [19], [20]. Additionally, the machine-readable RoPA can support organisations in the mandatory demonstration of compliance during a regulator inspection, a certification audit, or a standardised certification accountability framework (see Art. 42). The ability to seamlessly exchange RoPA information with regulators and other compliance stakeholders, such as certification bodies, also requires the machine-readable RoPA to be interoperable. This opens new possibilities for automated or semi-automated validation of accountability for external certifications, seals, and codes of conduct.

Despite these potential benefits and the importance of a standardised interoperable RoPA, only a limited number of RoPA-related explorations have been conducted in academic research, where existing efforts are limited to early-stage work involving enterprise architecture models [21]. Though several research approaches have been using Semantic Web standards for GDPR compliance in general (see Section 2.5), few have addressed the RoPA.

1.2 Research Question

The research question for this thesis is: '*To what extent can an electronic approach to RoPA support the demonstration of compliance with the Accountability Principle of the GDPR?'*

For clarity, each term within the research question is defined and discussed below:

- *An electronic RoPA approach (the ERoPA Approach)* — A machine-readable approach to gathering, presenting, transferring, and reviewing information to support the implementation of RoPA.
- *Support the demonstration of compliance* - providing appropriate records and measures [22] as set out in Article 5 (1) of the GDPR (see Section 2.3.1).
- The *Accountability Principle of the GDPR* (see Section 2.2)—The GDPR Art.5 Accountability Principle states that controllers are responsible for and must be able to demonstrate compliance with the other six principles of data protection: (i) Lawfulness, Fairness, and Transparency, (ii) Purpose Limitation, (iii) Data Minimization, (iv) Accuracy, (v) Storage Limitation, (vi) Integrity and Confidentiality (see Art. 5). This means controllers must ensure they comply with the principles and have appropriate processes and records to demonstrate compliance. For this thesis, the notion of accountability is limited to the use of records to demonstrate compliance, and the actual verification of compliance or establishment of processes required for compliance is not in scope of this work.

1.3 Research Sub Questions

The research question is answered through the following four research sub questions (RSQ):

RSQ1: What are the stakeholders' requirements for the ERoPA Approach?

RSQ2: How to create an interoperable specification for implementing the ERoPA Approach by answering the following sub-questions:

RSQ2a: What information is required to be maintained for the ERoPA Approach?

RSQ2b: How can this information be represented as a Semantic Web ontology for the machine-readable and interoperable representation of information required by the ERoPA Approach?

RSQ2c: How can the information required by the ERoPA Approach be communicated between stakeholders?

RSQ3: To what extent does the ERoPA Approach support implementing GDPR accountability?

RSQ4: What are the key considerations for organisations implementing an ERoPA Approach?

Each research sub question will be addressed in a dedicated chapter of the thesis. Please refer to Table 1 for the chapters where each RSQ is addressed.

Table 1 Overview of Which Thesis Chapters Address Each Research Sub Question

Chapter	Chapter Title	Research Sub Question Addressed
2	State of the Art	RSQ1
3	Research Methodology	
4	Requirements for the ERoPA Approach	RSQ1, RSQ2a
5	Design and Evaluation of the Semantic Model of RoPA	RSQ2b
6	Design and Evaluation of a Specification for RoPA Interoperability	RSQ2c
7	The ERoPA Approach and Upsilon Case Study	RSQ3, RSQ4
8	Conclusions	All

1.4 Methodology

The Action Design Research (ADR) methodology [1] was selected to develop the ERoPA Approach artefact, which consists of a machine-readable ontology for representing RoPA, an interoperability specification to enable transfer of ERoPA data, tools and methods for validation and quality assurance control and validation via SHACL (Shapes Constraint Language), and SPARQL query for retrieving required information, and deployment guidelines to enable an organisation to configure tools and processes to take advantage of the ERoPA Approach (see Section 4.2). This approach supports the creation of an optimised ERoPA Approach artefact through the interaction of design efforts and contextual factors throughout the development of the ERoPA Approach components: i) the common semantic model of RoPAs (CSM-RoPA) (see Section 5.1), ii)_the Semantic RoPA Interoperability specification named the Data Protection Catalogue (DPCat) (see Section 6.1) and iii) the ERoPA deployment guidelines. The ADR methodology is an iterative process, beginning with an initial problem formulation stage. This is followed by three building, intervention, and evaluation (BIE) stages with researchers, practitioners and users iteratively contributing to the ERoPA Approach artefact. These stages are not one-time events but a continuous cycle, with researchers, practitioners, and users contributing to these BIE stages. The process concludes with a guided emergence of the optimal design, leading to a formalisation of learning [1]. The ADR research stages [1] for the ERoPA Approach artefact are presented in Figure 1.

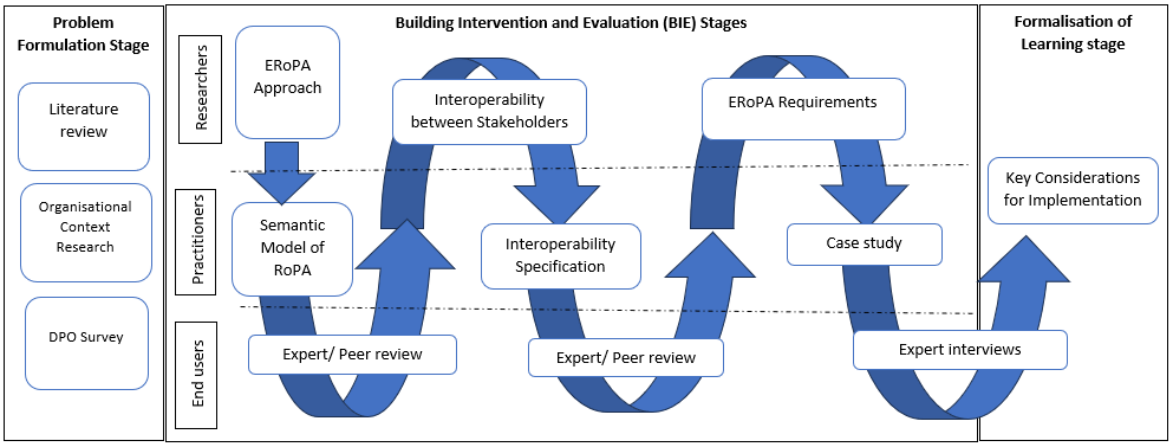

Figure 1 Action Design Research Methodology Approach Overview

The ADR methodology requires the designation of mutually influential roles such as researchers, practitioners, and end users (see Table 2 for the definition and allocation of each role). For this research, the author of this thesis, who is a practising data protection officer[4] is the 'researcher' who formulates ideas manifested in design (e.g. the ERoPA concept originating from GDPR RegTech). The role of practitioner is carried out by i) the organisation's data protection team members, who conduct the planning and implementation of the ERoPA for the organisation; and ii) the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) members. The DPVCG is made up of technologists and legal experts whose focus is the development of the Data Privacy Vocabulary (DPV); their expertise is leveraged to also support the development of the ERoPA Approach.

Table 2 Definition and Allocation of Action Design Research Roles.

ADR Role	**Definition**	**Who will conduct this role?**
Researcher	A researcher conducts systematic and organised investigations into the subject matter.	The author of this thesis is a practising Data Protection Officer and a prototype developer.
Practitioner	Information system practitioners are involved with planning and implementing IT resources for their organisations.	The Data Protection team from the case study organisation, and the DPVCG,
End User	An end user is a person who uses a product or service.	Data Protection Experts, Peer reviewers, Conference attendees, and industry users of ERoPA components.

[4] The author is a certified Data Protection Officer, working in professional practice.

Several groups function as 'end users' for this research. These include external data protection experts, academic peer reviewers, conference attendees, and industrial users of ERoPA components. The perspectives and contributions of each of these end users are gathered to contribute to the optimal design for ERoPA.

1.5 Technical Approach

Ontology Engineering: The technical approach for developing the ontologies used the NeOn methodology [23], a comprehensive framework for the collaborative development of networked ontologies. A NeOn requirement specification was prepared to determine the ontology's scope, purpose, intended use, and intended users. The non-functional requirements are identified, and the competency questions that the ontology (based upon the GDPR Article 30 RoPA requirements) should support were documented. These steps consisted of gathering data from GDPR Article 30 (see Section 4.5.1) and regulator-supplied RoPA templates (see Section 4.5.2) to identify the information requirements and develop ontological concepts [13]. A systematic review of the concepts was conducted to establish synonyms, overlapping concepts, and related concepts using Scharffe's mapping classifications [24]. Based on the interpretation of the GDPR and the use of concepts in ROPA, direct relationships were made, such as composition or qualifications, and establishing domain and range that were implicit in the templates.

The ontologies were designed using OWL [25] and the Protégé ontology editor [26] by employing a bottom-up approach which extended suitable existing work from the State of the Art (see Section 2.5.4) through a manual process that matched GDPR concepts with existing ontologies. Three iterations of matching cycles were conducted from 2020 to 2024 to iterative develop the ontologies and extend it with advances in the State of the Art. The Ontologies were documented with human-readable annotations where each class and property had labels added as per WIDOCO [27] best practices, which ensures that the ontology adheres to standards for usability and clarity.

Storage and Collaboration: For managing the technical artefact and working with collaborators, GitHub[5] was used as a shared repository for resources, including specifications, use cases, and test data. Google Drive was used to store survey results.

1.6 Evaluation Approach

This section describes the evaluations completed for each of the ADR BIE loops through the development of the ERoPA Approach's CSM-RoPA in the first BIE stage, CSM-RoPA plus DPCat

[5] GitHub: Let's build from here · GitHub

Interoperability specification in the second BIE stage and the ERoPA deployment case study with all ERoPA Approach components in the third BIE cycle.

The **evaluation of the CSM-RoPA model** sought to establish (i) whether the ontology contains the necessary concepts and relationships to represent a GDPR RoPA, and (ii) whether the CSM-RoPA ontology meets Semantic Web standards and best practices based on FAIR principles for research data (findable, accessible, interoperable and reusable)[28]. The evaluation consists of five parts: (i) an analysis of the extent to which CSM-RoPA ontology meets the competence questions set out in the ontology requirements specification (See Section 5.2.2), (ii) a review of the extent to which the CSM-RoPA ontology meets Semantic Web best practices, (iii) a use-case of how CSM-RoPA can express a RoPA, (iv) the presentation of CSM-RoPA to the DPVCG peer review, and (v) an overview of the peer-reviewed publications and industry use of CSM-RoPA.

The **evaluation of the DPCat interoperability specification** determined the extent to which DPCat meets the Interoperability specification competence questions. The evaluation includes a case study involving five common GDPR scenarios to gather and transfer GDPR RoPA compliance information among stakeholders based on the established use cases (U1-U5). To demonstrate the capability of DPCat to meet these competencies and use cases, a series of real-world 'typical' DPO tasks, taken from the Data Protection Professionals survey (see Section 4.6.2), are conducted using DPCat.

The **ERoPA case study deployment evaluation** was conducted in a real-world organisation, which is referred to as Upsilon. The case study consisted of three evidence bases collected and synthesised to gather key findings. The first data collection was direct observations gathered during the case study prototype deployment, following Morgan's framework [29] (see Section 7.4.2). Once the deployment observations were gathered, they were synthesised based on common patterns identified and presented as a summarised set of observations which underpin a set of deployment guidelines for EROPA.

The second data collection consisted of semi-structured interviews with data protection experts, using a prepared video that provided an overview of the ERoPA deployment for the case study, which were analysed using pattern analysis. The third data collection in the case study validated a real-world use-case for its extent to satisfy the UK GDPR regulator (ICO) Accountability requirements. The findings from the case study (based on a synthesis of the three data collections) were combined with the state-of-the-art analysis and the Data Protection Professionals survey to provide guidelines for organisations considering ERoPA deployment. These were synthesised in a Zachman Framework to formalise the learning stage of the ADR methodology and support organisations in implementing the ERoPA Approach in their own use-cases [30].

1.7 Contributions

The development of the ERoPA Approach to RoPA maintenance and interoperability has resulted in one major contribution and one minor contribution.

The major contribution is the ERoPA Approach, which (i) formulates the requirements for a machine-readable RoPA for GDPR based on stakeholder needs (ii) meets these requirements with the CSM-RoPA ontology for RoPA representation and (iii) the DPCat RoPA interoperability specification for exchanging information between stakeholders (iv) provides tools and methods such as RDF conversion, a data catalogue for metadata management, quality assurance control and validation via SHACL (Shapes Constraint Language), and SPARQL query for retrieving required information and (v) deployment guidelines for organisations considering the ERoPA Approach.

The ERoPA Approach offers data controllers a comprehensive metadata foundation for creating a machine-readable RoPA that assists Data Protection Officers (DPOs) in meeting GDPR accountability requirements and identifying gaps in compliance. DPCat enables stakeholders to exchange GDPR accountability information efficiently, leading to a consistent and accurate RoPA across organisations. Both approaches surpass conventional methods that use spreadsheets in terms of information management and potential for automating GDPR compliance, and enable the establishment of communication channels for compliance information amongst stakeholders – a key concept in the EU's vision of Data Spaces [31].

The research also delivers **a minor contribution** as the improvement of the Data Privacy Vocabulary (DPV), which is a state-of-the-art resource for representing GDPR information using semantic web standards, where the findings of this thesis were contributed to improve DPV's coverage of GDPR and RoPA information through membership and participation in the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG).

This research has resulted in 9 peer-reviewed scientific publications, including 3 journal publications, 5 conference papers, and 1 book chapter, which are listed chronologically below:

1. Ryan, P.; Crane, M. and, R. (2020). *Design Challenges for GDPR RegTech*. In Proceedings of the 22nd International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-423-7; ISSN 2184-4992, pages 787-795. DOI: 10.5220/0009464507870795 [32].
2. Ryan P., Crane M., Brennan R. (2021) *GDPR Compliance Tools: Best Practice from RegTech. In: Enterprise Information Systems*. ICEIS 2020. Lecture Notes in Business Information Processing, vol 417. Springer, Cham. https://doi.org/10.1007/978-3-030-75418-1_41 [17].

3. Ryan, P., Pandit, H., Brennan, R.: *A Common Semantic Model of the GDPR Register of Processing Activities* (2020), Legal Knowledge and Information Systems Conference, Jurix 2020 doi:10.3233/FAIA200876 [13].

4. Ryan, P. and Brennan, R. (2021). *Demonstrating GDPR Accountability with CSM-RoPA: Extensions to the Data Privacy Vocabulary*. In Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-509-8; ISSN 2184-4992, pages 591-600. DOI: 10.5220/0010390505910600 [33].

5. Ryan, Paul, Pandit, Harshvardhan J., & Brennan, Rob. (2021, June 20). *Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV*. International Conference on Semantic Systems (SEMANTiCS), Amsterdam, Netherlands. DOI 10.3233/SSW210043 [34].

6. Ryan, P., Brennan, R. *Support for Enhanced GDPR Accountability with the Common Semantic Model for ROPA (CSM-RoPA)*. SN COMPUT. SCI. **3**, 224 (2022). https://doi.org/10.1007/s42979-022-01099-9 [19].

7. Ryan, P.; Brennan, R.; Pandit, H.J. *DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR*. Information **2022**, 13, 244. https://doi.org/10.3390/info13050244 [18].

8. Pandit, H. J., Esteves, B., Krog, G. P., Ryan, P., Golpayegani, D., & Flake, J. (2024). *Data Privacy Vocabulary (DPV) -- Version 2.0* Springer Nature https://doi.org/10.1007/978-3-031-77847-6_10 [6].

9. How to Manage My Data? With Machine--Interpretable GDPR Rights! https://doi.org/10.3233/FAIA241254 [35].

This research has generated 2 components of the artefact, as presented in Table 3:

Table 3 Artefact Component Parts Produced as part of this Research

ID	Component Description	Sub-component	Target Research Sub-Questions
C1	Requirements list for the ERoPA Approach		RSQ1 / RSQ2a
C2	The ERoPA Approach	Semantic model of RoPA (CSM-RoPA)	R02b / RSQ3
C3		RoPA Interoperability Specification (DPCat)	RSQ2c / RSQ3
C4		Guidelines to assist organisations deploying the ERoPA Approach	RSQ4

The development of the ERoPA Approach has also had several impacts on the field. From the regulatory institution perspective, the ERoPA Approach and associated research has been presented to the European Union Data Policy and Innovation Unit (CNECT.G.1)[6] which is responsible for the policies and enforcement of GDPR, and to the Jamaican data protection regulatory authorities, who have utilised this approach to regulatory compliance by implementing an automated solution for working with RoPA for their GDPR-inspired data protection law. The ERoPA Approach was also presented at the Science Foundation Ireland (SFI) Summit 2022, which is the primary research funding agency in Ireland (now called Research Ireland), and at two Academic-industry events within the ADAPT Research Centre. Within industry, the ERoPA Approach has been utilised by Signatu [36] for developing automated solutions for GDPR compliance, while Alias Law[7] has expressed interest in implementing and carrying out additional research on the ERoPA Approach. The work also led to the establishment of a new standards development project in ISO/IEC JTC1 SC27 WG5 committee on Information security, cybersecurity and privacy protection for "PII Processing Records" based on the GDPR's ROPA [37].

1.8 Thesis Overview

The rest of this thesis is structured as follows. Chapter 2 provides a state-of-the-art overview of GDPR RoPA compliance in organisations. This chapter addresses **RSQ1**. It provides (i) an overview of the Accountability Principle under the GDPR, (ii) an overview of RoPA, and the integral role that it plays in demonstrating GDPR compliance and the importance of RoPA in supporting the DPO in carrying out their duties (iii) current organisational approaches to RoPA compliance, and provides an overview of the strengths and weaknesses of these organisational approaches, (iv) technological

[6] https://op.europa.eu/en/web/who-is-who/organization/-/organization/CNECT/COM_CRF_37139
[7] https://www.olympe.legal/

approaches to RoPA, and the opportunities that GDPR RegTech offers for GDPR compliance particularly with semantic interoperability of heterogeneous data (v) the chapter concludes by identifying the critical requirements of a GDPR machine-readable RoPA: the ability to automatically collect, represent, transfer, and review heterogeneous GDPR RoPA information between RoPA stakeholders.

Chapter 3, the methodology chapter, provides information to support the reader and an overview of the research approach followed in the thesis. Chapter 4 compiles the requirements for ERoPA (**RSQ1**). These requirements were derived from examining regulatory publications, reports and guidance documents, and insights from academic research on RoPA and GDPR compliance automation. Additionally, they reflect commercial practices gathered from a survey conducted with 42 Data Protection Officers (DPOs). These requirements are presented as a specification for ERoPA. The chapter also addresses **RSQ2a** to identify the information required to be maintained in an ERoPA. This information was gathered through an analysis of seventeen regulator-supplied RoPA templates.

Chapter 5 describes the development of an ERoPA ontology to implement the Common Semantic model of RoPA (CSM-RoPA). The NeOn methodology [38] was utilised based on Semantic Web standards and best practices to address **RSQ2b.** The chapter details the ontology's design process, implementation, and evaluation. CSM-RoPA was developed from the GDPR concepts identified in RoPA templates gathered from data protection regulators (see Section 4.5) and modelled to express the semantic requirements of all RoPA templates. The evaluation establishes the extent to which CSM-RoPA meets i) the competency questions for a semantic RoPA, ii) follows ontology engineering best practices, iii) enables the representation of a RoPA and iv) gathers feedback from the DPVCG and peer-reviewed publications.

Chapter 6 builds on the previous chapter's common semantic model of RoPA. This chapter addresses **RSQ2c**. The chapter details the design, implementation, and evaluation of the Data Processing Catalogue (DPCat) specification for the interoperability and communication of RoPA between data protection stakeholders. DPCat was evaluated using a case study where GDPR accountability information was exchanged between stakeholders in four different scenarios to establish the extent to which DPCat could meet the interoperability specification competence questions.

Chapter 7 addresses **RSQ3** to establish the extent to which the ERoPA Approach supports the implementation of GDPR accountability. The Chapter presents a validating case study of deploying an ERoPA prototype in a real organisation using the CSM-RoPA ontology, the DPCat interoperability specification, tools such as a triple store, quality assurance specifications and query and reporting tool. The case study demonstrates how ERoPA can support typical DPO tasks. The ERoPA Approach is evaluated using (i) observations gathered through deployment, (ii) structured

interviews, and (iii) an assessment of the extent to support a regulator's accountability framework. The chapter also addresses **RSQ4** to identify the key considerations for organisations implementing an ERoPA Approach.

Chapter 8 presents conclusions and future work. It confirms that the ERoPA Approach can support GDPR RoPA accountability to the extent that the GDPR information is held within RoPA. In addition, the ERoPA Approach provides the DPO with automated tools and methods for performing typical DPO tasks. The guidelines for ERoPA deployment support the deployment of ERoPA. Still, the organisation would require a mature data governance capability to deploy it effectively. It would require the semantic modelling of other GDPR documents, such as data processing agreements and privacy notices, which extend beyond the RoPA domain to gain full GDPR accountability compliance.

2 State of the Art

2.1 Chapter Overview

This chapter provides an overview of the state-of-the-art for GDPR RoPA compliance with organisations. It addresses **RSQ1** to determine what are the stakeholder requirements for the ERoPA Approach. This chapter is broken into five sections as follows:

1. An overview of the Accountability Principle under the GDPR (see Section 2.2).
2. An overview of the RoPA under the GDPR, the integral role that RoPA plays in demonstrating GDPR compliance and the importance of RoPA in supporting the DPO in performing their duties (see Section 2.3)
3. A review of current organisational approaches to RoPA compliance and provides an overview of the strengths and weaknesses of these organisational approaches (see Section 2.4).
4. A review of technological approaches to RoPA and the opportunities that GDPR RegTech offers for GDPR compliance, particularly with semantic interoperability of heterogeneous data (see Section 2.5)
5. A synthesis establishing the critical requirements of a GDPR machine-readable RoPA (see Section 4.7).

2.2 An Overview of the GDPR Accountability Principle

This section discusses the GDPR Accountability Principle and its obligations on organisations. It also provides an overview of the methods, such as accountability frameworks and GDPR certification, that organisations use to demonstrate GDPR compliance.

Accountability is a common principle for organisations across many disciplines; it requires that organisations live up to expectations, for instance, in delivering their products and behaviour towards those they interact with [39]. The Anglo-Saxon word 'Accountability' has a broadly understood meaning of how responsibility is exercised and how it is made verifiable [40]. It can be viewed as an expression of how an organisation displays 'a sense of responsibility—a willingness to act in a transparent, fair and equitable way' and 'the obligation to explain and justify conduct [39].

The GDPR applies these definitions by placing accountability as one of the seven fundamental principles of the regulation and requires that an organisation be responsible for and demonstrate compliance with all principles of the GDPR [22]. These inform and permeate all other

provisions of that legislation. They are at the heart of every organisation's approach to processing personal data [22]. The Accountability Principle of the GDPR mandates organisations to ensure 'appropriate and effective measures to put into effect the principles and obligations of the GDPR' as provided by GDPR in Art.42. Organisations must demonstrate compliance with the GDPR upon request from the supervisory authority. The GDPR obligates organisations to establish whether, how, and how well they protect personal data. Considering such a significant obligation, organisations must fundamentally rethink how they store and process personal data organisation wide [41].

The purpose of accountability is not just the evaluation of compliance with statutory obligations [42]. An accountable organisation can go beyond the minimum legal compliance requirements to demonstrate how they respect the privacy of their data subjects, i.e. the subjects of the processing of personal data [42]. The organisation has several audiences for the demonstration of compliance [33]. Internally, it must demonstrate that it is operating accountable to its corporate board and employees and put internal organisational privacy and information management programs in place. The GDPR Accountability Principle includes implementing internal measures and procedures, putting into effect existing data protection principles, ensuring their effectiveness and the obligation to prove this should data protection authorities request it [43].

Similarly, the organisation has obligations to demonstrate compliance to external stakeholders such as members of the public, business partners, shareholders, and civil society bodies representing individuals, as well as to Data Protection Authorities [42]. The organisation may also need to demonstrate compliance with a certification body as part of a code of conduct or a standardised certification accountability framework provided by GDPR in Art.42. External certifications, seals, and codes of conduct can support accountability when accompanied by external validation. In this way, verification and demonstration are ensured [44], [45].

Accountability gives organisations a solid framework for compliance with applicable legal requirements, protecting data subjects from harm to privacy, and building trust in the organisations' ability to engage in the responsible use of data [42] . Importantly, accountability provides an approach to data protection that is transparent, risk-based, technology-neutral and future-proof [42]. Implementation of accountability increases trust in the organisation's operations [42]. It ensures that the organisation is equipped to manage new challenges to data protection law and practice, regardless of advances in technology or changes in the behaviours or expectations of individuals [42]. It also gives them flexibility and agility in customising their data privacy management programs. Successfully embedding accountability will enhance the business's reputation of being trusted with personal data [46].

The alternative to implementing a robust accountability framework is that the organisation may face the consequences for non-compliance with the Accountability Principle of the GDPR. The

implications for such non-compliance can result in an organisation facing fines up to €20 million, or up to 4% of the annual worldwide turnover of the preceding financial year, whichever is greater (see Art.83). Hence, the GDPR Accountability Principle can be viewed as a double-edged sword, where one side contains the reputational trust and confidence gained from acting accountable when organisations are meeting their obligations versus compliance failures resulting in reputational and financial damages.

2.3 The Record of Processing Activities (RoPA) under the GDPR

The GDPR requires organisations acting as a data controller to maintain a RoPA detailing all processing activities under their responsibility (see Art.30). A data controller is an entity that determines the purposes and means of processing personal data. Similarly, data processors (entities that process data on behalf of the controller) must also record all processing activities conducted on behalf of each data controller they work with (see Art.30). The RoPA, as a measure to demonstrate compliance, is part of how data controllers demonstrate and implement the principle of accountability (see Art.5.2) [11]. A well-drafted RoPA will demonstrate to a regulator that a data controller is aware of and has considered the purpose of all processing activities and risks occurring within the organisation [11],[47]. The RoPA should assist the data controller in demonstrating that the organisation has considered the risks and implications of processing personal data, the specific and limited personal data required for each activity, and the appropriate technical and organisational measures to secure the personal data to be processed. The GDPR requires that organisations maintain a RoPA and make it available to the data protection regulator upon request [11].

2.3.1 The RoPA as an Integral Component of GDPR Accountability

The GDPR requires all controllers and processors to maintain a RoPA unless an organisation employs fewer than 250 persons (see Art.30). The obligation to complete a RoPA is mandatory for any organisation where the processing is likely to result in a risk to the rights and freedoms of data subjects, and the processing is not occasional, or the processing includes special categories of data or personal data relating to criminal convictions and offences is required to maintain a RoPA regardless of its size (see Art.30).

RoPA is an integral element in demonstrating GDPR compliance [10], [11], [33], [48]. Creating a RoPA requires gathering GDPR accountability information from multiple stakeholders. This information may be collected from organisational units or external entities such as processors and may take many forms (such as privacy notices and data processing agreements, for example). Supplementing the RoPA with complementary accountability information can make it an accurate

control tool for compliance with the GDPR [48]. The role that stakeholders play in contributing such GDPR accountability information to RoPA is critical, where multiple stakeholders each have integral roles in the RoPA. Figure 2 shows how organisational units need to collect GDPR information from internal sources (e.g. Privacy Notices and Data Processing Agreements) and external sources (e.g. processors). This data is submitted to and represented in the RoPA. The DPO and external entities such as auditors, certification bodies and regulators then review the RoPA. The role of each stakeholder is described in more detail in Section 2.3.3.

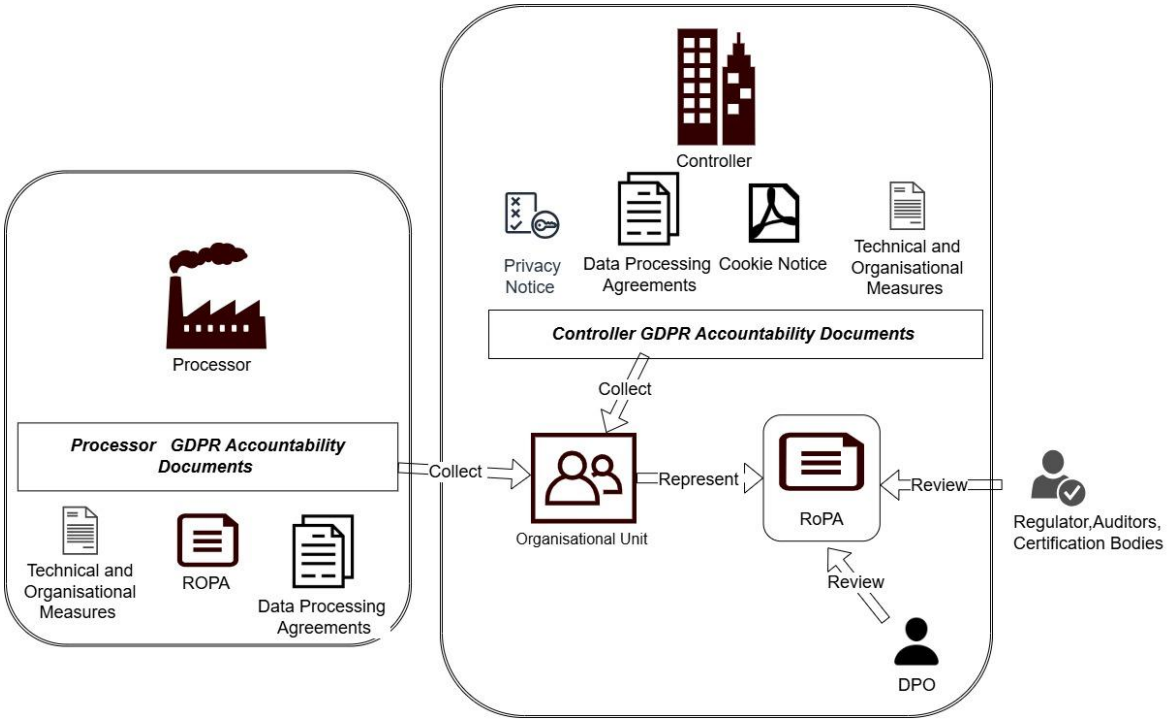

Figure 2 Stakeholder Contributions of GDPR Accountability Documents to RoPA.

The gathering of data for RoPA may be a challenging process, but a well-drafted RoPA will explain to the GDPR regulator that a data controller is aware of and has considered the purpose of all processing activities within the organisation [11], [47].

2.3.2 Information Present in a RoPA

GDPR Article 30 prescribes the information that the RoPA must contain and states that controllers and processors must be able to provide such records to a regulator on request [11], [49]. Article 30 states, 'Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility 'and 'the record shall contain all of the following information' as follows:

A. the name and contact details of the controller and, where applicable, the joint controller, the controller's representative, and the data protection officer;

B. the purposes of the processing;
C. a description of the categories of data subjects and the categories of personal data;
D. the categories of recipients to whom the personal data have been or will be disclosed, including recipients in third countries or international organisations;
E. where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers, the documentation of suitable safeguards (see Art.49.1);
F. where possible, the envisaged time limits for erasure of the different categories of data;
G. where possible, a general description of the technical and organisational security measures (see Art. 32)

Whilst Article 30 of the GDPR specifies what data must be maintained within a RoPA, Regulators may look for additional information to be included. For example, regulatory guidance from the Irish Data Protection Commission (DPC) advises that the RoPA should include relevant extra information as 'helpful extras' [11]. A sample RoPA is presented in Figure 3 [11]. The DPC's guidance advises that *the RoPA must be easily understood*. This will help if different business areas and employees input information into the RoPA so that it will be evident to a new reader which information is mandatory, and which data has been added as an extra [11]. The DPC suggest that RoPA should also contain additional data beyond the Article 30 requirements as follows:

- The Article 6 legal basis for processing.
- The Article 9 basis for the processing of special category data.
- Whether a breach has occurred in respect of a specific processing activity.
- The data transfer legal basis relied upon for third-country transfers.
- Risk level the organisation may have assigned to each processing activity.

While Article 30 clearly states the mandatory regulatory requirements for a RoPA, further analysis of regulator-provided RoPA templates shows that many regulators suggest supplementary data to be included in RoPA. These additional requirements are detailed in Chapter 4.

Approach A: An example of a well completed RoPA

A Company Record of Processing Activities

Data Controller: A Company Ltd
123 A Street
Dublin
info@acompany.ii
DPO: David P Other

GREEN COLUMNS ARE **MANDATORY**

BLUE COLUMNS ARE AS **APPLICABLE**

Reference	Processing Activity	Sub Processing Activity	Process Owner	Purpose of Processing	Categories of Data Subject	Categories of Personal Data	Data Accessed by	Joint Controller	3rd Country transfer	International Organisation Transfer	Safeguards for transfer	Retention	Technical and Organisational Security Measures	Legal Basis	Legislation	Risk Register Reference	Notes
1.1	Staff Payroll	Reconcile Hours Worked	HR	Pay Calculation and Working Time requirements	Staff	Name Staff number Clocking hours	HR	NA	No	NA	NA	7 Years (Financial Records)	Password protected system Encrypted archive after 1 year	6(1)(b) Contract 6(1)(c) Legal Obligation	Working Time Act, 1997	NA	Electronic clocking system Accessed by DS and their line manager for day-to-day admin (See Process 15)
1.2	Staff Payroll	Staff Pay	HR	Issue Staff Pay	Staff	Name Staff Number Address PPSN Bank Details Length of Service Sick status Deductions	HR	NA	No	NA	NA	7 Years (Financial Records)	Password protected system Encrypted archive after 1 year	6(1)(b) Contract		NA	
1.3	Staff Payroll	Staff Pay	Accounts	Instruct Bank to Pay Staff	Staff	Name Address Bank account details	Accounts HR ABC Bank	NA	No	NA	NA	7 Years (Financial Records)	End-to-end encryption	6(1)(b) Contract		6.6	Low risk of bank compromise of personal data

Figure 3 An Example of a RoPA Provided by the Data Protection Commission [11]

2.3.3 Stakeholder Roles regarding RoPA.

Many stakeholders partake in completing the collection, representation, transfer, review, and inspection of a RoPA. This section describes each of the stakeholders and the roles they play within RoPA management.

Organisation Units (OU) are base unit subdivisions that group resources, typically for management and administrative purposes, e.g. Human Resources or Sales Team [11]. Organisational units allow organisations to efficiently structure and manage resources, applying specific policies, permissions, and configurations to each OU as needed. Organisational units are vital to collecting and gathering the organisation's processing for inclusion in RoPA as they conduct the actual processing, so they know what data is processed, what data processors are engaged in, and what technologies are employed.

A **Data Controller** is an individual or organisation that determines the purposes, conditions, and means of processing personal data (see Art.4.7). Data Controllers are the entities responsible for demonstrating GDPR compliance. They must ensure that all personal data processing activities are recorded on RoPA and available for presentation to a regulator. This requires the gathering of RoPA information for stakeholders, the recording of this information and the sharing of this information with regulators and auditors. Controllers (and Data Processors) appoint **Data Protection Officers** (DPO) who, in many cases, coordinate the gathering and recording of processing activities. The DPOs use the RoPA to review processing activities for the purposes of identifying the risks for flagging to the controller. The DPO often acts as the liaison between regulators, auditors, and certification bodies and arrange that any requests for compliance information required to demonstrate compliance are met.

A **Data Processor** is an individual or organisation that processes personal data on behalf of a Data Controller [50] (see Art. 4.8). The data processor follows the data controller's instructions but does not have independent authority to decide the purpose or means of the data processing. The role of the data processor with RoPA is that the processor must maintain their own RoPA, detailing the processing that they are conducting on behalf of a controller.

The **regulator**'s role regarding RoPA is that they can request a copy of a ROPA to be presented on demand.

Certification bodies and auditors' roles are like those of regulators, where data is exchanged with such parties to monitor compliance.

2.3.4 RoPA and the Role of the Data Protection Officer (DPO)

Many organisations appoint a Data Protection Officer (DPO) to assist with meeting their GDPR compliance obligations[10], [51] . To conduct the role, the DPO requires a broad set of skills in GDPR legal compliance, and a detailed knowledge of the organisation's business processes [51]. The DPO collaborates with stakeholders, such as data subjects, employees, processors, and regulators and provides consultancy and guidance on business processes (see Figure 2). The role involves a broad spectrum of activities, from maintaining a RoPA to dealing with data breaches to completing data protection impact assessments. The DPO must have visibility of all activities and monitor and report compliance to the highest level in the organisation [52]. The DPO is 'privacy on the ground' in that the DPO is the early warning system for GDPR compliance within the organisation [53].

For the DPO, accessing a comprehensive record of the organisation's data processing operation is a primary step for GDPR compliance [49]. The DPO must conduct an in-depth review of all the personal data processing operations to see whether they comply with the GDPR[49]. The overall picture of information assets, personal data, and the related processing operation provided by the RoPA is the first step to accountability since it enables the evaluation of risk on the rights and freedom of individuals and the implementation of appropriate technical and organisational measures to ensure a level of security appropriate to the risk [49].

The RoPA provides the information for the DPO to assess the risks posed by the processing activity [47]. This information within RoPA helps the controller to '[take] into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons' posed by each personal data processing operation, and to 'implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed by this Regulation' (see Art.25). A clear and documented understanding of a company's data processing activities provides executives with crucial information on actual risks, allowing them to make informed decisions about which risks they are willing to accept and which they are not [47].

The RoPA is a crucial input for privacy processes, and many privacy processes can be linked to and leveraged upon the RoPA [54]. One example is using RoPA to help respond to a data access request, where the organisation can quickly identify the right source of information in the RoPA. Another example is completing a privacy notice; the RoPA is vital for achieving such documents [54]. Similarly, when conducting a Data Protection Impact Assessment on an existing process, one can reuse existing information from RoPA and assess only the incremental risk instead of starting from nothing. Hence, the RoPA is a vital record that assists the DPO in performing its duties to monitor GDPR compliance.

2.4 Current Organisational Approaches to GDPR Compliance

This section provides an overview of the tools and methods available to organisations to support their GDPR compliance programs. [17]. These will include (i) manual and approaches, (ii) enterprise software solutions, (iii) maturity models, (iv) accountability frameworks, and (v) GDPR certification. These categories of tools have been identified based on categories taken from research from the IAPP [12] and previous publications from the researchers [32], [33].

2.4.1 Manual and Informal Approaches to GDPR Compliance

The introduction of the GDPR precipitated the need for a substantial amount of GDPR compliance documents to be prepared and stored by data controllers as a form of evidence to demonstrate accountability [55]. Many regulators have curated manual templates to assist organisations in recording their GDPR accountability evidence [13], [18]. Regulators have provided DPIA templates [56], RoPA templates[18], and self-assessment checklists such as the DPC 'GDPR readiness checklist tool' [57] to enable organisations to check their GDPR accountability status. Many private privacy consultants and industry bodies also provided templates for GDPR accountability documents [58]. This led to organisations adopting a manual approach to GDPR accountability in the early days of GDPR enforcement [12].

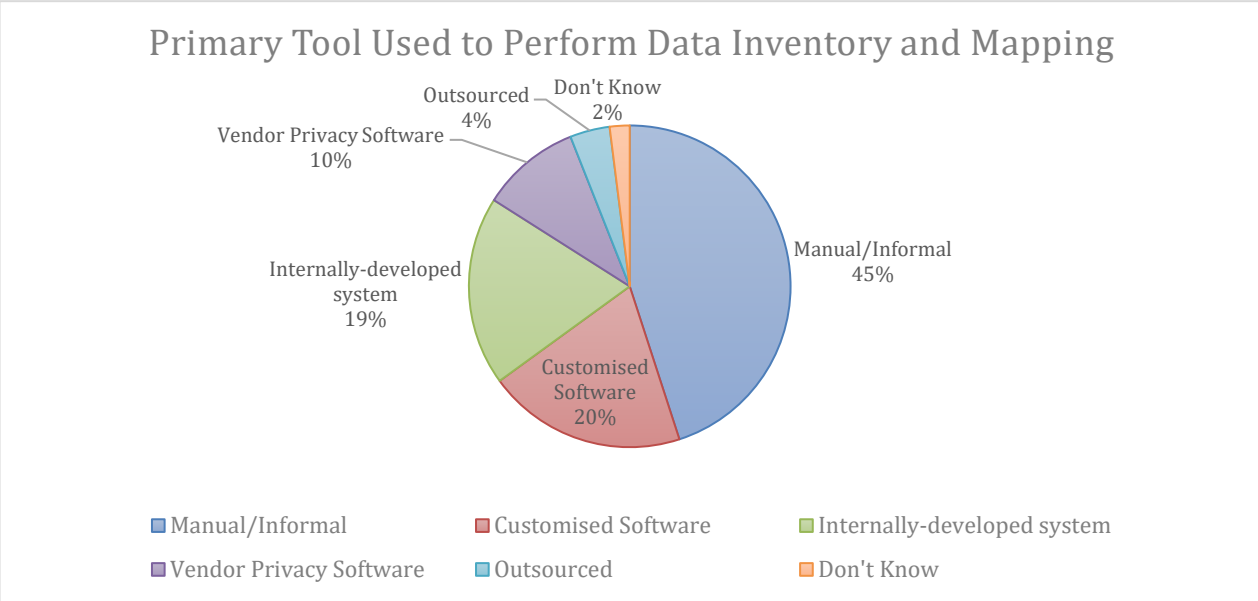

Figure 4 Primary Tools for Data Inventory and Mapping (IAPP) [12]

In 2019, the International Association of Privacy Professionals (IAPP) examined GDPR compliance practices [12] and found that almost half (45%) of organisations completed data mapping and inventory operations using manual or informal tools, such as spreadsheets, email, and in-person

communication. An extract of the IAPP report is presented in Figure 4. Approximately 10% of organisations utilised vendor-supplied software off the shelf [12]. The IAPP's recent survey from 2022 indicates a move away from manual spreadsheet solutions, which corresponds with a growth in the domain of privacy management software and solutions. However, the landscape for privacy solutions is still largely fragmented, with organisations taking a wide range of approaches that are often proprietary and non-interoperable as presented in. These approaches are presented in Figure 5 for an extract taken from the IAPP survey report[12].

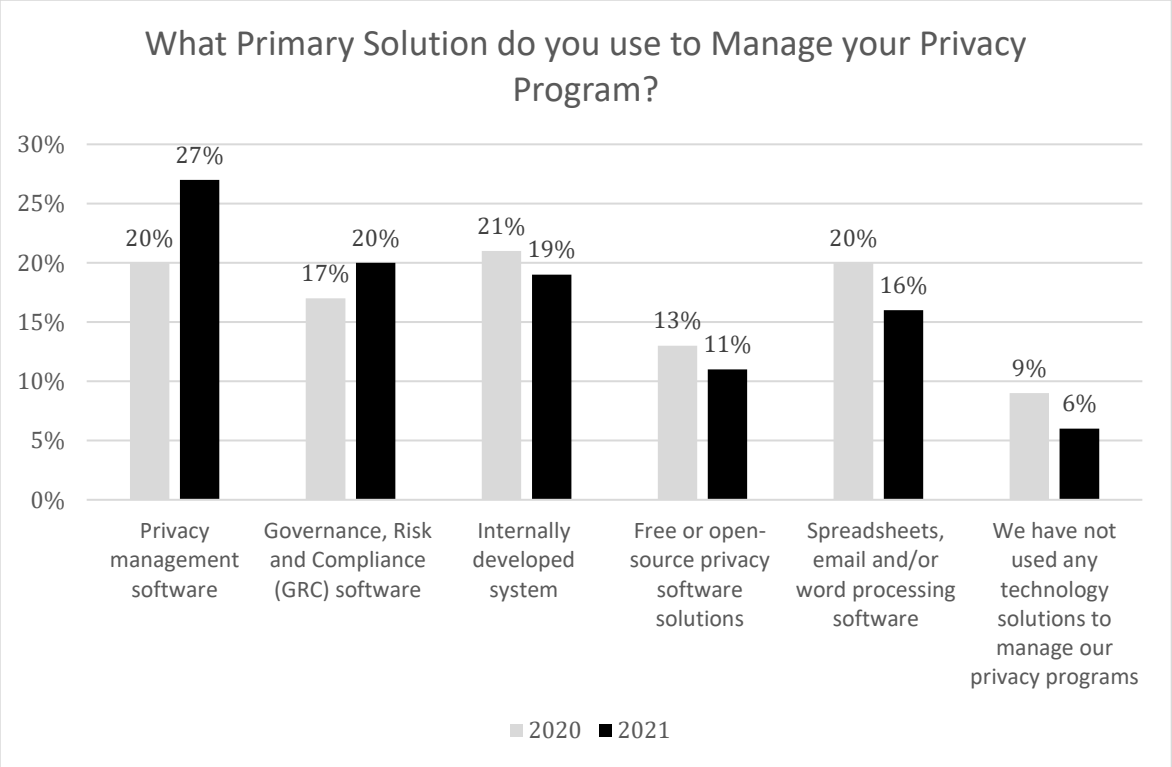

Figure 5 Primary Solution Used to Manage Privacy Program in 2020-21 (IAPP) [19]

A manual approach to GDPR compliance, supported by manual regulatory templates, has many disadvantages. Among them are that they are difficult to maintain, they are labour intensive, and they lack Interoperability with other systems [17], [59].

2.4.2 Enterprise Software Solutions

There has been a noted need for tools and methods to assist organisations in meeting their GDPR compliance obligations [12]. This need is being met by significant financial investments by venture capital companies, with over $500 million invested in privacy-related start-ups worldwide in 2017 [18]. This has continued in 2020 with significant funding going into privacy software vendors, such

as One Trust[8] $210m, Ave Point[9] $200m, Privitar[10] $80m and BigID[11] $50m [60]. In 2024 the global privacy software market was forecasted from $3.8 billion in 2024 to $48 billion by 2032[61] . As of 2022, there were 364 vendors offering privacy software tools to organisations [62]. This had grown from 2017, when there were forty-four such vendors and compliance solutions around data protection and privacy regulations (see Figure 6).

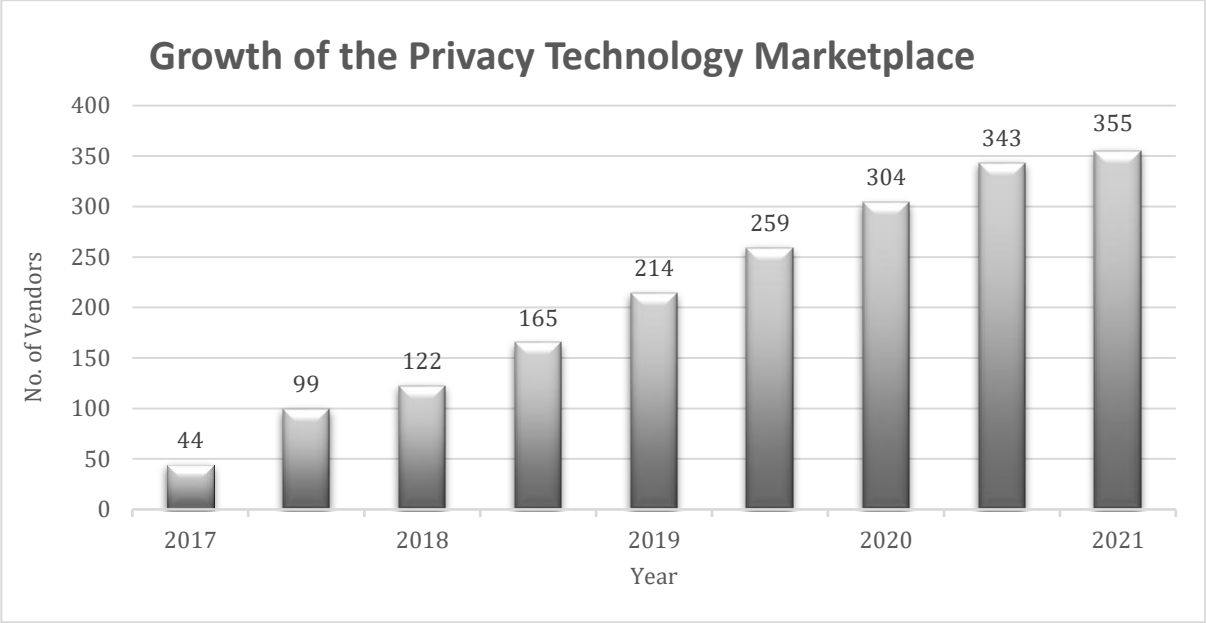

Figure 6 Growth of Privacy Technology Marketplace [62]

These software solutions come in many forms (see Table 4), from simple questionnaires and templates to solutions that focus on individual compliance with GDPR, such as statistical discovery tools for Data Subject Access Requests (DSARs), where individuals exercise their right to information under Article 15 of the GDPR. The main categories of these privacy tools are as follows [17], [62]:

- Activity Management – control and monitor access to personal data.
- Assessment Managers - automate different functions of a privacy program, locating risk gaps, demonstrating compliance.
- Consent managers - help collect, track, demonstrate and manage users' consent.
- Data discovery – determine and identify personal data held.
- Data mapping solutions - determine data flows throughout the enterprise.
- Data Subject Access Requests (DSARs) – automation of the process
- De-identification pseudonymisation tools

[8] https://www.onetrust.com/
[9] https://www.avepoint.com/
[10] https://responsum.eu/
[11] https://bigid.com/

- Secure internal enterprise communications
- Data breach incident response solutions
- Privacy information managers - provide the latest privacy laws around the world.
- Website scanning – catalogue cookies

While vendors offer a variety of privacy software solutions, as Table 4 displays, 'there is no single vendor that will automatically make an organisation GDPR compliant' [18]. Most solutions offered by private enterprise solution providers cover three or fewer categories (see Table 4, which combines data from the IAPP Privacy Tech Vendors 2020 and 2021) [60], [63].

Table 4 Privacy Software Tools, Number of Vendors per Category [60], [63] .

Privacy Product Category	**No. of Vendors offering this service in 2019**	**No. of Vendors offering this service in 2020**	**Increase 2019 to 2020**	**% Increase 2019 vs 2020**
Activity Monitoring	86	93	+7	8.1%
Assessment Manager	102	118	+16	15.7%
Consent Manager	82	94	+12	14.6%
Data Discovery	91	107	+16	17.6%
Data Mapping	114	140	+26	22.8%
Data Subject requests	-	53	+53	Not available in 2019
De Identification/Pseudonymity	45	55	+10	22.2%
Enterprise Communications	39	54	+15	38.5%
Incident Response	62	86	+24	38.7%
Privacy Information Manager	72	96	+24	33.3%
Website Scanning	30	37	+7	23.3%
Total	248	328	+80	32.3%

Many organisations conduct their data mapping process for RoPA using a manual approach [12]. Data mapping involves identifying what personal data the organisation holds and how the information flows through an organisation [49]. There are 140 privacy software vendors that offer data mapping solutions for this activity [63]. Additionally, there are examples of vendors that provide mapping solutions that allow for auto-generation of their RoPA, such as Mandatly[12] and

[12] https://mandatly.com/

Mineos[13]. The benefits of the data mapping process are that it provides visibility into hidden or unknown systems and helps increase the completeness of RoPA. While such data discovery tools benefit the DPO, vendor-supplied software does not meet best practices as they are mainly proprietary, not standards-based, and not interoperable [19], [64]. Many privacy vendors offer APIs or integrations to overcome these challenges, which come with additional development costs. Hence, an organisation employing such privacy software can face significant costs in successfully interoperating with third parties or even intra-company. Signatu provides a novel chatbot interface to create a RoPA entry based on inputs provided by a user [36]. These approaches enable the creation of RoPA but face challenges like those of other data protection vendors, with a need for interoperability with third parties.

The key benefits of Privacy Software, as claimed by the most prominent vendor, One Trust[14] are as follows:

- Streamline data protection compliance management activities.
- Prevent the use of out-of-date or incorrect information in critical decisions.
- Reduce time spent locating information assets.
- Accurately measure data protection risks and privacy program maturity
- Remove manual data discovery processes, reducing cost and improving output.
- Demonstrate progress easily with automated data protection compliance reporting.
- Improve productivity with automated workflows, pre-built templates, and integrations.
- Real-time insights and analytics to demonstrate data protection compliance.
- Translate complex data protection and compliance requirements into practical deliverables.
- Faster completion of RoPA with collaboration and integrations into productivity tools

Privacy software solutions go some way towards helping the organisation with a demonstration of compliance; research has identified several weaknesses in private enterprise software solutions[17], [32], as follows:

- They are not supported by published methodologies, they are not supported by standards, and they lack evidence to support their validity or even utility.
- Many of these solutions are stand-alone in that they lack interoperability with other GDPR compliance systems and hence cannot easily be assembled into tool chains providing comprehensive compliance reports and metrics, quality improvement processes or data analytics such as root cause analysis.

[13] https://www.mineos.ai/

[14] https://www.onetrust.com/solutions/gdpr-compliance/

- They focus on manual or semi-automated assessment approaches that are labour-intensive, rely on domain experts and are not driven by qualitative operational data that organisations are increasingly generating.
- They are created by private enterprises based on interpreting the regulations rather than being developed with the regulator's input.

These solutions offer an organisation a starting point for GDPR compliance. However, the lack of academic rigour or formal regulatory input and the inability to connect and build toolchains inhibits these solutions.

2.4.3 Maturity Models

Capability Maturity Models have been used for privacy compliance monitoring since 2011[17], [65]. The American Institute of Certified Public Accountants privacy maturity model [65] was used to understand an organisation's privacy compliance standing using a set of questions called 'generally accepted privacy principles' in 73 measurable criteria. It gauged compliance along five maturity levels from ad hoc to optimised. The drawbacks of this methodology as a measure of compliance with the GDPR are that it predates the GDPR and would, therefore, need updating to reflect the new regulation. The more recent International Association of Privacy Professionals (IAPP) Maturity Framework [54] develops a series of checklists built through 'collaboration between a team of highly experienced privacy and security professionals, lawyers and regulators.' The French GDPR Regulator CNIL provided a self-assessment [66] that describes the five levels of maturity (ad hoc, repeatable, defined, managed and optimised) for the eight following items: define and implement privacy procedures; pilot privacy governance; maintain data inventory; ensure legal compliance; training and awareness; handle data subjects requests; manage security risks; and manage data breaches. For instance, a maturity level of five is reached when data inventories are used as a steering tool for privacy activities. All three solutions provide visualisations of compliance on an axis and are an indicative measure of compliance. However, they do have several drawbacks [17], [32]:

1. they are labour-intensive and dependent on highly skilled labour (domain experts)
2. they are infrequently updated.
3. they are typically not automated, or do not form part of an automated process toolchain.
4. they lack standardised reporting mechanisms, especially machine-readable ones.

While these maturity models serve as indicators of an organisation's GDPR compliance, the limitations outlined hinder their potential for further deployment as part of enterprise information systems. However, with future automation or the adoption of suitable reporting standards, these tools can overcome their lack of reporting and interoperability standards, paving the way for enhanced GDPR compliance.

2.4.4 Accountability Frameworks

This section provides an overview of how Accountability Frameworks are used for GDPR compliance. The researcher has identified two such frameworks based on his professional experience as a DPO and a Google search using the term “GDPR Accountability Framework”. The two frameworks identified are: (i) the Centre for Information Policy Leadership (CIPL) framework and (ii) the ICO accountability framework (see Section 3.2).

In 2018, the Centre for Information Policy Leadership (CIPL) [42] developed its accountability-based data privacy and governance programs, which provide more definition of the key elements of accountability in a privacy context, as shown in Figure 7. These key elements are Leadership and Oversight, Policies and procedures, Training and Awareness, Response and Enforcement, Transparency, Monitoring and Verification, Risk Assessment, and Monitoring and Verification.

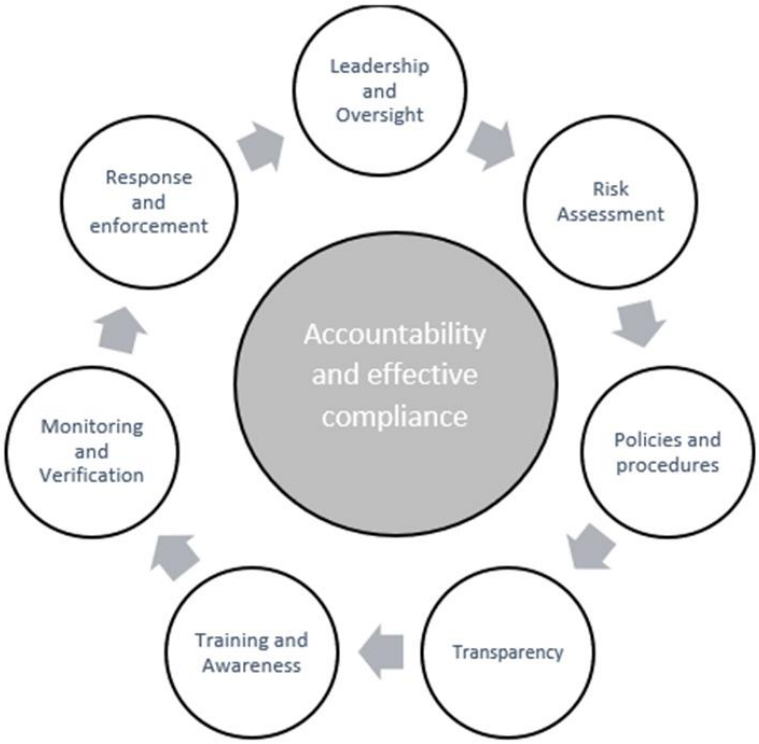

Figure 7 The Accountability Wheel/Universal Elements of Accountability from CIPL [42]

The enforcement of the GDPR in 2018 provided significant challenges to organisations in attaining compliance with its requirements [67]. Many regulators provided guidance documents on specific GDPR topics to help organisations understand their compliance obligations [11], [48]. The UK Information Commissioner's Office (ICO) published its GDPR accountability framework in 2020 to support organisations in gathering a comprehensive overview of their compliance across the breadth of the GDPR [68]. The framework was developed in 2020 when the United Kingdom was subject to the EU GDPR, and it has remained unchanged since its inception; hence, it still aligns with

the EU GDPR. The ICO describes this framework as a tool for organisations, large or small, to assess the effectiveness of the accountability measures they have in place and understand where they need to improve [68].

Several data protection supervisory authorities have provided self-assessment checklists and accountability toolkits to assist organisations in preparing for GDPR. Their purpose is to support organisations in assessing whether they have appropriate and effective internal policies, procedures, and measures to ensure compliance with data protection regulations. These come in the form of a series of questions and checklists designed to assist the organisation in checking its compliance level. These toolkits are devised to provide broad coverage of all the principles of the GDPR. A mature example of such a framework is the ICO accountability framework (see Section 3.2). The framework, which consists of ten compliance areas, sets out expectations for how organisations can demonstrate accountability. Each expectation is further detailed in multiple statements, and organisations must assess their compliance with each statement on a four-level scale. A regulator's provision of an accountability tracker is a progressive step in describing GDPR accountability, as it provides clear expectations that organisations must meet.

Just like maturity models, these checklists provide an overview of compliance. However, the main drawbacks of these tools for GDPR compliance are as follows[17], [32]:

- they are fundamentally high-level self-assessment tools, generic by nature, need more details to be applied to automated tools and are focused on organisational rather than technical factors.
- they rely on users' qualitative input and lack input or output interoperability with other solutions.

However, the key benefit of these checklists and toolkits is that they have been developed with the input of regulators, unlike maturity models and private enterprise software solutions, which have been developed independently.

2.4.5 GDPR Certification

GDPR certification allows organisations to demonstrate compliance measures to individuals, other organisations, and supervisory authorities. It encourages organisations to comply with GDPR provisions and enables them to be certified for having appropriate safeguards for personal data processing [69]. Certification provides a public-facing accountability tool for demonstrating compliance measures and benefits to data subjects by allowing them to assess the level of data protection an organisation offers [45], [69], [70]. Certification schemes specify the mechanisms for processing personal data and how controls and measures are implemented. Accredited bodies

assess and validate organisations; certified organisations are reviewed every three years to ensure compliance.

Certifications completed by approved independent third parties can offer significant benefits to organisations [69] . The certification process builds confidence in data subjects and requires organisations to gather evidence of GDPR accountability, which can be manual and costly, particularly for smaller organisations [17]. It's important to note that certification alone does not guarantee GDPR compliance, and additional measures may be required [71]. The prominent challenge organisations face with certification schemes is that gathering data and evidence for certification is primarily manual. While there are six approved GDPR Article 42 certification bodies[72], to date, the researcher has been unable to find any evidence that any of these six accreditation bodies offer an automated compliance solution [72].

Certification provides meaningful opportunities to demonstrate GDPR compliance. However, the tools required to collect a wide range of GDPR accountability heterogeneous data are currently unavailable. The ability to gather data from different sources and seamlessly integrate with certification systems would efficiently demonstrate GDPR compliance and eliminate the labour-intensive barriers associated with certification.

2.4.6 Overview of Existing Organisational Approaches to GDPR Compliance

This section examines five organisational approaches to GDPR compliance, which are (i) manual and informal approaches, (ii) enterprise software solutions (iii) maturity models (iv) accountability frameworks and (v) GDPR certification.

The guidance from the DPC shows that organisations face significant challenges related to GDPR RoPA compliance (see Section 1.1). The key challenges faced by organisations are failing to maintain accurate and up-to-date RoPA containing the necessary detailed information [11]. While many organisations still use spreadsheets for RoPA compliance, it is argued that the move to privacy software solutions has only partially addressed their challenges. Table 5 presents an analysis of existing GDPR compliance RoPA approaches based on the RegTech capabilities and success factors [15], [16], [17], [32], [64], [73]. The table examines each of the existing approaches based upon (i) manual effort to maintain (ii) interoperability with other systems, (iii) capable of handling Heterogeneous data sources, and (iv) machine-readable capability. The analysis identified that manual approaches require significant effort to maintain the RoPA. Furthermore, existing approaches lack machine readability and interoperability with other systems containing GDPR-relevant information, although some enterprise software solutions partially address this. Similarly, existing approaches partially address the ability to manage data from heterogeneous data sources.

Evidence from Data Protection Officers (DPOs) highlights the challenges they face due to a lack of human resources in terms of compliance with current data protection legislation [74]. DPOs also report that up to 50% of DPOs are part-time (less than 20% FTE[15]) from the Norwegian data protection regulator [74].

Table 5 A Comparison of the Capabilities for Existing GDPR Compliance Approaches.

GDPR Compliance Approach	**Manual effort to maintain**	**Interoperability with other systems**	**Capable of handling Heterogeneous data sources**	**Machine-readable capability**
Manual	High	No	No	No
Enterprise Software Solutions	Medium	Partially	Partially	Partially
Maturity models	High	No	No	No
Accountability Frameworks	High	No	No	No
GDPR certification	High	No	No	No

The analysis of the challenges faced by existing approaches to GDPR RoPA compliance indicates that three key areas need to be addressed to overcome existing challenges. The first is the need to **represent all GDPR accountability data comprehensively**. This requires GDPR information in all systems to use a common vocabulary to express GDPR concepts.

This information must be available in a machine-readable format supporting compliance automation. Maintaining an accurate and up-to-date RoPA requires exchanging GDPR information between stakeholders, and **interoperability** must be in place to enable this transfer between organisational units, DPO and regulatory bodies. Currently, the exchange of GDPR accountability documentation relies heavily on manual processes or predefined Application Programming Interface (API) solutions, which are vendor-specific and limited to integration with that specific privacy vendor software. The challenges faced by Data Protection Officers (DPOs) highlights the need for an efficient approach to GDPR compliance.

[15] Full time equivalent

2.5 Technological Approaches to GDPR Compliance

This section provides an overview of the emerging technical approaches that offer opportunities for automating GDPR compliance. It looks at the emergence of RegTech, the factors that have contributed to its success, and how these might be applied to GDPR compliance. The section also examines technical initiatives for GDPR compliance, particularly regarding technical approaches to RoPA.

2.5.1 The Success Factors of RegTech

RegTech can be defined as 'the use of technological solutions to facilitate compliance with and the monitoring of regulatory requirements' [73] . RegTech has played an essential role in making regulatory compliance more efficient and effective [15]. RegTech is an information technology (IT) that (a) helps firms manage regulatory requirements and compliance imperatives by identifying the impacts of regulatory provisions on business models, products and services, functional activities, policies, operational procedures and controls; (b) enables compliant business systems and data; (c) helps control and manage regulatory, financial and non-financial risks; and (d) performs regulatory compliance reporting [15]. RegTech has evolved to address regulatory challenges in the financial system through innovative technology. It can support the technical handling of large amounts of data, sophisticated analysis and automated data processing within intermediaries and between intermediaries and supervisors [16] .

The first key feature of RegTech is the adoption of new technologies by financial institutions. The Fintech revolution of 2008 has been a critical contributor to the success of RegTech [75]. The evolution of Fintech has seen rapid growth and the creation of new opportunities through the application of Big Data, Internet of Things, Artificial Intelligence, machine learning, distributed ledger technology and blockchain, smart contracts and digital identity [16]. The advent of regulations such as Know-Your-Customer (KYC) and Anti-Money Laundering (AML) [76] and EU directives such as Markets in Financial Instruments (MiFID and MiFID II) [77] has meant that organisations are required to invest in technology to meet their compliance obligations. These complex regulations require automated compliance monitoring and reporting. Organisations can no longer rely on humans to monitor and report complex areas like insider trading and money laundering[17]. Technology has enabled the automatic detection and reporting of irregular activities to a compliance officer, thus reducing the risk of human error in the form of an inattentive staff member[33], [73].

Similarly, RegTech helps organisations determine what level of investment advice must be given to a customer based on the results of an automatically processed questionnaire [73]. Again, this solution helps an organisation to reduce errors and meet its legal obligations through process

automation. These solutions reduce the need for human intervention and make compliance less complex. RegTech tools are used to leverage data from existing operational information systems and seek to provide agile solutions to improve compliance visibility through the automation of mundane compliance tasks and reduction of risk to the organisation [78]. The foundation of compliance has been to prevent, identify, respond to and remedy risk [17], [78]. RegTech solutions enhance compliance basics through improved data integration, automation, predictive analytics and strategic process alignment [17], [78].

The second key feature of RegTech systems has been the digital transformation of data [16]. The KYC regulation requires an organisation to gather and validate information relating to their customers. Traditionally, this was a complex manual process requiring the presentation of documents, recent photographs and proof of identity and address. The digitisation of KYC data has dramatically reduced the risk of errors occurring [16]. Similarly, the use of automation and AI to monitor AML regulations for the purposes of regular surveillance of transactions has improved regulatory compliance with the prevailing AML guidelines [79] . Distributed ledger technology (DLT) has enabled smooth and seamless information-sharing between financial organisations and regulators [80]. This has resulted in the faster delivery and verification of KYC data and significant savings in time spent on the manual reconciliation of documents [80]. Upward of 80% of enterprise data today needs to be more structured [81]. Unstructured data resides in emails, files, PDFs, or documents that do not have a pre-defined data model. The digital transformation of such data has facilitated the adaption of new technologies and enabled the organisation to automate processes, gain efficiencies and optimise regulatory compliance [82] .

The third key feature of RegTech systems is the application of agreed semantic standards through the ability to create machine-readable meta-models that enable data virtualisation across heterogeneous data stores [15]. Data in siloed, heterogeneous databases can be virtualised, and ontologies, predictive analytics/machine learning algorithms, and AI may then be applied to classify the data automatically. It can then be used to monitor compliance, advise on risk, and inform stakeholders. When such a knowledge-based model is linked and integrated with a regulatory knowledge base such as legal texts, it is possible to enable automated regulatory compliance reporting of regulations as part of a business process [15]. This allows an enhanced data-driven approach to the management risk. According to Butler et al. - 'Semantic interoperability ensures that these exchanges make sense— that the requester and the provider have a common understanding of the 'meanings' of the requested services and data' [15]. RegTech holds much promise for regulators and organisations. The full benefits of RegTech will only materialise if the pitfalls of a fragmented approach are avoided. Semantic standards are the key to all this [15].

The fourth key feature of the success of RegTech has been the role of the supervisory authority as an enabler has played, both as a beneficiary and an enabler of RegTech systems. In

2014, Andy Haldane, then Chief Economist at the Bank of England, said, 'I have a dream. It is futuristic but realistic. It involves a 'Star Trek' chair and a bank of monitors. It would involve tracking the global flow of funds in close to real-time, such as with global weather systems and internet traffic. Its centrepiece would be a global map of financial flows, charting spillovers and correlations' [83].

Table 6 Features of RegTech Systems

Publication	**Adoption of developments in technology**	**Digital Transformation of data**	**Common standards and Agreed semantics, Interoperability of systems**	**Role of supervisory authority as enabler/ Stakeholder**
Fintech, RegTech and the reconceptualisation of Financial Regulation [83]	✓	✓		✓
Fintech and RegTech Impact on regulators and banks [84]				✓
Fintech and RegTech in a nutshell and the future in a sandbox [85]				✓
Adaptive financial regulation and RegTech, a concept article on realistic protection for victims of bank failures [86]	✓			✓
RegTech compliance and the technology judgment rule [87]	✓			✓
The emergence of RegTech 2.0 from Know Your Customer to Know Your Data [75]	✓	✓		✓
An innovative RegTech approach to financial risk monitoring and supervisory reporting [80]	✓	✓	✓	
RegTech is the New Black – The Growth of RegTech Demand and Investment [88]	✓			
Understanding RegTech for digital regulatory compliance [15]	✓	✓	✓	✓
RegTech as a response to regulatory expansion in the financial sector [73]	✓	✓	✓	✓
Fintech and RegTech: Enabling Innovation while Preserving Financial Stability [15]				✓
The road to RegTech is an astonishing example of the European Union [16]	✓	✓		✓

The regulator now has access to periodic or real-time, fine-grained compliance reports and incremental improvements in compliance but also promotes the design of a regulatory framework able to adapt to new rules and regulations dynamically [83]. The regulator seeks tool development in areas of interest such as real-time and system-embedded compliance/risk evaluation tools, big data techniques, visualisation and automation tools, software integration tools and cloud technologies. Tech Sprints have been conducted since 2016 to prove that standards-based RegTech

could help automate and make the task of regulatory reporting by financial institutions more efficient and cost-effective [15]. The Tech Sprint proposed solutions using Natural Language Processing, human and machine-readable rules, machine-executable regulatory reports, and linked knowledge bases.

This literature review has identified the four critical success factors of RegTech systems: (1) the adaption of technological developments, (2) the digital transformation of data, (3) agreement on common standards to facilitate interoperability of systems, and (4) the role of regulators as facilitators for the automation of regulation (see Table 6). The regulator's role as an enabler is featured in all literature, and it was the primary success feature for RegTech. Without the regulator's facilitation to make regulations digital, RegTech could not have happened [15]. Adopting new technologies is widely seen as a critical success factor and a key enabler to the success of RegTech. The role played by the digital transformation of data, the need for common standards and agreed semantics, and the interoperability of systems are less pronounced in the literature but are still seen as critical elements in building RegTech systems.

2.5.2 RegTech for Supporting GDPR Compliance

This section examines the current research literature regarding technological approaches to GDPR compliance (GDPR RegTech). The four critical success features of RegTech systems identified in section 3.5.1 are used to evaluate how these features could be applied to improve GDPR compliance. The RegTech approach to GDPR outlined here was first proposed by the author in 2020 as a set of design challenges and published in the 22nd International Conference on Enterprise Information Systems [32]. Each success factor is discussed to establish how they could be applied in a GDPR context to facilitate GDPR compliance.

An additional Google Scholar search was conducted in 2022 to identify the top publications using the search term 'GDPR Compliance Automation Technology.' The scope of this search was to identify publications that discuss compliance with the entirety of the GDPR rather than specific GDPR compliance areas such as privacy notices or consent. This search yielded twenty-three publications. These twenty-three publications were reviewed to establish the main approaches and evaluate the extent of the application of RegTech's success factors in these academic publications. The results of this analysis are presented in Table 7. The study showed that all academic GDPR RegTech approaches use new technologies such as Artificial Intelligence (AI), Natural language processing (NLP), Machine Learning, and Semantic Web to enable automated GDPR compliance.

Table 7 GDPR Compliance Automation - Analysis of the Primary Approach Taken

Publication Name	**Primary Technical Approach**	**Publication Contains these RegTech Success Factors**			
		Adoption of new technologies	Digital transformation of data	Agreement on Common Standards	Importance of Regulatory support
GDPR Compliance Tools Best Practice from RegTech [17]	Semantic Web Ontology	✓	✓	✓	✓
Exploring automated GDPR-compliance in requirements engineering: A systematic mapping study [89]	Natural Language Processing	✓	✓		
GDPR Compliance in the Context of Continuous Integration [90]	Compliance Testing Model				
Compliance Generation for Privacy Documents under GDPR: A Roadmap for Implementing Automation and Machine Learning [91]	Machine Learning	✓			
Privacy, Security, Legal and Technology Acceptance Elicited and Consolidated Requirements for a GDPR Compliance Platform [92]	Data Governance Platform		✓		
NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR [14]	NLP, Semantic Web	✓			
Artificial Intelligence-enabled Automation for Compliance Checking against GDPR [93]	NLP, ML	✓			
Automated GDPR compliance assessment for cross-border personal data transfers in Android applications [94]	ML	✓			
A Combined Rule-Based and Machine Learning Approach for Automated GDPR Compliance Checking [95]	Natural Language Processing, ML				
GDPR Compliance Verification in the Internet of Things [96]	Blockchain contracts	✓	✓		
Data Protection by Design Tool for Automated GDPR Compliance	Semantic Ontology	✓	✓		

Publication Name	Primary Technical Approach	Publication Contains these RegTech Success Factors			
		Adoption of new technologies	Digital transformation of data	Agreement on Common Standards	Importance of Regulatory support
Verification Based on Semantically Modeled Informed Consent [97]					
Towards Automated GDPR Compliance Checking [98]	NLP, ML	✓	✓		
Using Models to Enable Compliance Checking against the GDPR: An Experience Report [99]	UML representation of the GDPR	✓	✓		
Design Challenges for GDPR RegTech [32]	Semantic Web	✓	✓	✓	✓
Using Artificial Intelligence to Support Compliance with the General Data Protection Regulation [100]	Artificial Intelligence technologies: rule-based and machine learning	✓			
Automating GDPR Compliance using Policy Integrated Blockchain [101]	GDPR Ontology with Blockchain	✓			
Automating GDPR Compliance Verification for Cloud-hosted Services [102]	Business Process modelling	✓			
Monitoring the GDPR [103]	Monitoring of System Logs	✓			
On Enabling GDPR Compliance in Business Processes Through Data-Driven Solutions [104]	Process mining techniques	✓			
A Framework for GDPR Compliance for Small- and Medium-Sized Enterprises [105]	Manual compliance framework				
An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance [106]	Semantic Web technologies, Natural Language Processing	✓	✓		
A Method for Managing GDPR Compliance in Business Processes [107]	Business Process modelling	✓			

The first success factor for GDPR RegTech is *adopting new technologies* in the same manner that has been at the forefront of RegTech's successes. From the literature review of 'GDPR Compliance Automation Technology' publications a set of technical approaches to support compliance automation have been identified (Table 7). Five instances of Natural Language Processing are utilised for GDPR automation to meet compliance requirements such as checking Data Processing agreements [14]. Similarly, machine learning is being used to check compliance with international transfers [94], blockchain is being used to check contracts [96]; Business Process modelling has also been used to monitor GDPR compliance [107]; artificial intelligence is being used to check GDPR compliance [100]. Semantic Web ontology has been used in four GDPR compliance publications in Table 7. Semantic Web is discussed in more detail in Section 3.5.4.

The second success factor for GDPR RegTech is *the digital transformation of data*. This requires organisations to develop and build a dedicated data management capability [41]. Organisations face challenges meeting their GDPR obligations due to a lack of common ground between legal and data management domains. Labadie et al. propose dedicated data management capabilities within organisations to overcome this challenge [41]. This would act as 'an abstraction layer between the normative aspects of the regulation' to solve organisational problems [41].

Digitally transformed organisations need clearly defined data principles, where data is viewed as an asset [108]. The agreed uses of that data must be clearly defined, and the organisation must ensure that the use of data relates positively to the regulatory environment. They need to set out the organisational behaviours for data quality, who will access the data, how data will be interpreted (metadata), and how long it will be retained [109]. An organisation processing personal data must be able to locate and categorise all personal data (e.g. by using data catalogues plus data classification). The organisation must understand the level of sensitivity of the data processed, where it is stored, who is the data owner, who has access to the data, how data evolves through a lifecycle, what lifecycle stage it is at and how long the data is retained for, as it underlines the importance of data protection and privacy. Whilst the digital transformation of financial data in RegTech has facilitated the application of technology to this data, this may be more challenging in a GDPR environment where large amounts of data are held in manual formats [12].

The third success factor for GDPR RegTech is to make personal data interoperable between systems with the a*greement on common standards/ agreed semantics* for personal data processing. The semantic modelling of GDPR business processes could greatly benefit an organisation and provide machine-readable and interoperable representations of information that can be queried and verified based on open standards such as RDF, OWL, SPARQL, and SHACL [19], [110], [111] . When these models are combined with legal knowledge bases, they become beneficial for compliance evaluation and monitoring [17], [112]. This can help harmonise and facilitate a joint

approach between legal departments and stakeholders to identify feasible and compliant data protection and privacy regulation solutions [113]. There has been progress in developing 'Core Vocabularies', maintained by the Semantic Interoperability Community (SEMIC) [114], that provide a simplified, reusable and extensible data model for capturing fundamental characteristics of an entity in a context-neutral fashion in this area to foster interoperability. This work continues to be built on through the development of the W3C Data Privacy Vocabulary (DPV) [6] .

The fourth success factor for GDPR RegTech is the need for initiative-taking regulators who will work with organisations to automate regulation and make compliance more straightforward. The significant role of the regulator is found in only two publications [32] [17]. This was a significant success factor for RegTech but is given lesser priority in GDPR compliance automation literature to date.

2.5.3 Standard Initiatives to GDPR Compliance

The previous section evaluated the extent to which RegTech's best practices are being applied in emerging GDPR RegTech technologies. This section will discuss how these practices have progressed in GDPR compliance automation research and identify the areas where GDPR Regtech needs more progress to match the success of Financial RegTech.

While many approaches are being taken to automate GDPR compliance [14], [20], [98], [106], [115], each tends to be stand-alone research projects that focus on using a technology area to solve a particular compliance area of GDPR, such as Data Processing Agreements or Privacy notices. Whilst this is very positive, The GDPR requires organisations to demonstrate compliance across the full breadth of the regulation, so more holistic solutions are required.

An area of weakness with GDPR RegTech is the *digital transformation of data*. While the Fintech revolution facilitated this in the financial industry, the same cannot be said for personal data held by non-financial organisations [32], [116] . Effective digital data governance policies require machine-readable models of the organisation, its people and resources, its data processing activities and so forth. This means having a 'digital twin' (or data model) of the organisation and other relevant stakeholders, such as customers and the regulator [117]. Building these models is not enough; organisations also need data governance, management and processing platforms that can interact to track data, processing activities, and enforce policies.

Another significant weakness in GDPR RegTech is the *agreement on common standards/ agreed semantics.* Much of the research in compliance automation has focused on using ontologies that need to be interoperable with other ontologies or vocabularies. An agreed semantic ontology is required to represent all concepts in the GDPR. The DPV offers the potential to become such an ontology that can be customised [6]. The DPV has been developed between legal experts and technologists.

The literature review shows the benefit the *regulator's role as an enabler* for GDPR Compliance automation could bring. Data protection regulators have made some efforts to make compliance more accessible by providing guidance documents, self-assessment checklists, and templates [57]. Whilst each GDPR regulator must apply the GDPR consistently (GDPR recital 135), there is an urgent need for a more unified approach to technical solutions to facilitate GDPR compliance. An analysis of RoPA templates provided by regulators showed inconsistencies in templates, with some templates having as few as twelve fields, whereas others had forty-three fields [18] (see Section 4.5.2).

Several positive initiatives have been developed, such as the creation of the Internet Privacy Engineering Network (IPEN) [118], which promotes and advances state-of-the-art privacy engineering among regulators, academia, open source and business development, and other individuals committed to finding engineering solutions to privacy challenges. The objective is to integrate data protection and privacy into all development process stages, from requirements to production. The European Data Protection Board (EDPB) has also released open-source software known as the 'Website Evidence Collector' [119], which is a step in the right direction. There have been positive moves by the UK regulator (ICO) to develop sandboxes to work with innovators in privacy technology. However, this is very much in the early days. For GDPR RegTech to succeed, GDPR regulators need to move towards a symbiotic relationship with technology innovators and organisations processing personal data to develop open-source compliance tools, digital regulations, sandboxes [15] and tech sprints.

Enterprise Architecture (EA) may offer opportunities for organisations to conduct their own data mapping and catalogues to maintain RoPA [120]. Rozehnal proposed 'Enterprise Architecture' as an ideal source for representing processing activities and technologies in an organisation [120]. This is supported by Burmeister et al., who also investigated how enterprise architecture can provide a DPO with insights on organisational data processing activities concerning GDPR compliance [121]. Huth proposes using EA models, augmented with supplementary GDPR accountability information, to support the automated creation of a RoPA [21]. Huth's approach is that existing EA models can be enhanced with all relevant information necessary for creating a RoPA for reporting purposes. An advantage of EA models is that they can be in machine-readable format and can be used to store RoPA information. From the model, an up-to-date RoPA can be generated at any time. The EA RoPA approach may offer opportunities as an enabler for privacy compliance. EA modelling languages also suffer from heterogeneity and a lack of standardisation.

An early-stage ERoPA called the Universal record of processing activities (URoPA) has potential [122], [123]. It is a protocol designed to help companies and their DPOs comply with the legal requirements related to privacy laws worldwide. The URoPA initiative from Alias Law [122] has identified RoPA as the cornerstone of the data protection process. It states that companies rely

on static tools to maintain their RoPA (such as spreadsheets, as this form of RoPA does not reflect the reality of the company's processing. The Alias Law URoPA aims to address this issue with a universal and interoperable machine-readable record of processing. The twin goals of URoPA are to (i) Empower companies to automatically assess whether their policies are adequately implemented and detect anomalies whether they use a self-made or a third-party compliance tool and (ii) to enable interoperability between compliance tools and solutions so companies can centralise their privacy solutions without useless and time-consuming compatibility efforts. URoPA is modelled on the open application programming interface (API) specification [124] which can be combined loosely to achieve complex operations [115].

2.5.4 Semantic Web Approaches to GDPR Compliance

This section discusses the State of the Art of approaches utilising semantic web ontologies for RoPA, and broadly for GDPR compliance.

The **OntoRoPA Project** states that RoPAs should be more open and share information [115]. The OntoRoPA provides an ecosystem where metadata about RoPAs can be assessed with automatic and intelligent processes. It aims to create a standard ontology for RoPAs and to be able to share trustworthy and open information about RoPAs, ready to be exploited by intelligent processes in a community-based ecosystem. Linking high-quality data about RoPAs will allow for intelligent extraction of knowledge from these data protection items of information, flexible comparisons of RoPAs, and smart processes that assist the inspection of RoPAs.

The OntoRoPA uses ontologies, metadata, and blockchain to provide practical solutions for high-quality, trustworthy RoPAs. Successful Semantic Web approaches such as Linked Data and OWL are combined with blockchain technologies to ensure easy access, quality, and trust of RoPAs. OntoRoPA contributes new value with solutions for dependable, linked, high-quality semantics about RoPAs. In addition, it provides a model to evolve the ontology created by the research team to a standard supported by legal expertise.

OntoRoPA proposes the development of a domain ontology formally expressed in OWL that will be offered as open data that is reliable, reusable, and extensible. This professional ontology will support the creation and validation of RoPAs. Validation will be twofold: RDF validation for correctness and OWL validation for completeness. As stated, RoPA Ontology includes legal and professional knowledge extracted from the community of privacy and data protection experts—lawyers, legal advisors and scholars, data protection officers, and rulers proficient in creating and manipulating RoPAs.

Apart from OntoRoPA, no other approaches were found in literature that utilised semantic web ontologies for representing RoPA. However, as information for RoPA is itself a part of information required for GDPR, there exist several semantics-based projects that provide

ontologies, vocabularies, and policy languages that can be used to represent information for GDPR [6], [20], [97], [125]. **GDPRov** [126] , **GConsent** [127], and **SPECIAL** [128] provide ontologies for expressing GDPR-related concepts but do not incorporate RoPA requirements. **GDPRtEXT** [129] provides a vocabulary of GDPR concepts, of which some relate to RoPA (GDPR Art.30). **PrOnto** [130] provides concepts regarding data types, documents, agents and roles, purposes, legal bases (and more), but it is not available for reuse. These GDPR Semantic-based projects focus on legal compliance evaluation rather than deployment and interoperability[114] and often are limited to addressing a few rights and obligations [125] as a result of which they do not address or contain the concepts required for modelling RoPA.

BPR4GDPR (business process re-engineering and functional toolkit for GDPR compliance) is a relevant ontology-based compliance methodology that dictates and evaluates processes [131]. It is based on advanced process mining from event logs of IT systems to discover, monitor, and improve processes without pre-modelling them before mining. BPR4GDPR thus creates a novel process-monitoring architecture with constraints for conformance checking and automated evolution of workflows and processes to satisfy the rules. Significant trials and development will be needed before advanced techniques like this are widespread in conventional organisations. The key differences between RoPA and BPR4GDPR are that RoPA is a legal requirement requiring the documenting of data processing activities with a focus on compliance transparency, whereas BPR4GDPR is a technical solution focusing on the automation of process re-engineering rather than record keeping and compliance.

The **Data Privacy Vocabulary (DPV)**, developed by the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), enables the creation of machine-readable, interoperable, and standards-based representations for describing the processing of personal data. The DPV provides an ontology based on the GDPR, and further provides taxonomies for these concepts to aid practical use-cases in interoperable representation and exchange of information about personal data and its processing [6], [113]. The DPV specification represents an abstract model of concepts and relationships that can be implemented and applied using technologies appropriate to the use-case's requirements. This enables the expression of machine-readable metadata about using and processing personal data based on legislative requirements such as the GDPR. The DPV has several peer-reviewed machine-readable specifications for Data Breach [132], Consent [127] and Data Protection Impact Assessment [133], thus providing a peer-reviewed WC3 standardised approach. These semantic resources could be used for RoPA interoperability. The DPV can formalise the highest number of information flows and can represent the most informative items required by the GDPR compared to other GDPR semantic models [113]. The DPV fills a crucial niche in the State of the Art by providing a vocabulary that can be embedded and used alongside other existing

standards, such as W3C ODRL [134], which can be customised and extended for adapting to specifics of use cases or domains[6].

In efforts focused on integration with software development methodologies and infrastructures, two notable efforts are **TIRA**—an OpenAPI extension for REST architectures [135] and TILT, which provides concept integration within code for practical privacy engineering [136]. The German DPA effort to identify fields and requirements through a 'Standard Data Protection Model (SDM)' is also relevant here, given its focus on information systems [41]. These existing efforts, specifically the DPV, provide ontologies representing GDPR-relevant information in an interoperable manner and to utilise it for aggregation, querying, validation, and exporting information based on identified RoPA-related information flows using Semantic Web technologies.

Signatu is a private enterprise company that provides a novel chatbot interface to create a RoPA entry based on user inputs [36] . These approaches enable the creation of a processing specification that can be outputted as a RoPA. Signatu contributes to and uses elements of the Data Privacy Vocabulary (DPV). This differentiates from other developments as the DPV is built to a W3C standard. Signatu offers tools that enable RoPA records to be created from a cookie scan and output a privacy notice from their processing specification. Although Signatu provides many useful tools for the organisation, it faces challenges like those of other data protection vendors. It needs interoperability with third parties using a standardised agreed vocabulary.

2.5.5 Overview of Technical Approaches to RoPA Maintenance

Recognising the opportunity a technical approach to RoPA (based on GDPR RegTech) brings, Table 8 reviews all identified methods for maintaining RoPA. Each approach to RoPA is reviewed as follows:

- What Interoperability method is used for communications between stakeholders?
- Does this RoPA approach use regulator templates to create its ontology?
- Has this RoPA approach gathered Legal experts' input into its ontology?
- Does this RoPA approach use a W3C standard open vocabulary?
- Is this RoPA approach supported by peer-reviewed publications?
- Are there live implementations of this RoPA approach?
- Have the requirements of DPOs been gathered for this approach to RoPA?

Each RoPA approach is reviewed, and a summary of each of these approaches is in Table 8.

Table 8 Review of State-of-the-Art Approaches to RoPAs

Current Semantic Approach to maintaining RoPA	**Interoperability method**	**Based on regulator templates**	**Legal Experts input to ontology**	**Standards-based information**	**Peer Reviewed**	**Live Implementation**	**DPOs Requirements gathered**
Manual Spreadsheets	Limited exchange	✓				✓	
Custom in house Software	Limited exchange / API	✓	✓			✓	Partial
Privacy Vendor Software	Limited exchange / API	✓	✓			✓	Partial
Enterprise Architecture [21]	No interoperability				✓		
URoPA[16]	Custom Ontology		✓				
OntoRoPA [115]	Custom Ontology		✓		✓		
Signatu [36]	Proprietary /Semantic Web	✓	✓	Partial		✓	Partial

Table 8 shows there are *three distinct approaches to the interoperability* of processing data for RoPA. These approaches are manual, API-based and Semantic Web. An analysis of manual approaches to RoPA, find many live implementations but no use of open source, W3C ontologies, no peer reviews, and only partial engagement with DPOs to establish their requirements. A review of the semantic approach to RoPA shows the same lack of DPO engagement, no live implementations, and few academic peer reviews. Several semantic RoPAs that are in development offer distinct interoperability possibilities. There is recognition that a semantic machine-readable RoPA may offer opportunities to overcome the exchange of heterogeneous data with multiple stakeholders to alleviate organisations' challenges; however, current ERoPA initiatives need live implementation to prove the concept. Several semantic RoPAs are in development but untested in

[16] https://github.com/uRoPA-project/uRoPA

an industry setting. There is also a distinct need to establish DPOs' requirements for ERoPA. Although the ERoPA is being developed by researchers with data protection and legal experience, engaging with end product users to gather requirements when designing information technology artefacts is the best practice [1].

2.6 Conclusion

The chapter examines current organisational approaches to compliance with the GDPR Accountability Principle, the technological approaches to GDPR compliance, and semantic modelling as an approach towards GDPR compliance. The existing approaches to an automated RoPA that use ontologies offer opportunities to support GDPR compliance. However, the analysis has identified a distinct lack of an agreed semantic interoperability ontology and specifications to enable organisations to exchange GDPR RoPA accountability metadata to support the demonstration of compliance with the GDPR Accountability Principle. Some specifications have been offered for GDPR accountability for other GDPR compliance areas, such as data subject rights and data breaches from the DPV [6], [35], [132], however, these do not address RoPA. Current initiatives need an agreed approach to automating RoPA that would benefit all stakeholder groups. The analysis has verified that organisations are challenged with maintaining an accurate and up-to-date RoPA. The key lessons identified in this State-of-the-art review are as follows:

- There is significant potential for GDPR RegTech based on the four identified RegTech success factors, which are (i) the digital transformation of data, (ii) the adoption of new technologies (iii) an agreement on Common Standards and semantics (iv) an enabling regulator.
- Most semantic compliance initiatives have not studied semantic RoPA. There is no ontology for representing heterogenous RoPA information based on Semantic Web standards and best practices.
- There is no specification for the exchange of GDPR RoPA information, and this is required as both intro-organisation and inter-organisation information flows between tools are required for comprehensive RoPA documentation.
- The well-established DPV is an emerging standard ontology for all types of GDPR activities. The DPV already provides concepts for DPIA’s, Breach, and Consent. The DPV offers significant opportunities to develop a RoPA specification. , but before the work presented in this thesis, it lacked coverage of RoPA concepts.
- There is a lack of DPO and stakeholder input to semantic GDPR compliance technology.

- Tools for DPOs are immature and organisations are significantly challenged to meet the Accountability Principle of the GDPR despite significant investment and various approaches.
- The opportunity offered with standardised semantic RoPA where GDPR accountability data can be exchanged between stakeholders has not yet been realised.
- While many semantic RoPA initiatives are in the early stage of development, they lack maturity and need a common standardised ontology to enable interoperability.
- For a semantic RoPA to succeed, it must be based on an open semantic standard that fully expresses all the concepts found in the relevant RoPA, interoperate with RoPA relevant GDPR accountability document, and be possible for all stakeholders to use the RoPA.

This review has completed part of **RSQ1** and provided a new understanding of the landscape of GDPR RegTech for RoPAs that satisfy the Accountability Principle. The next chapter on ERoPA requirements formalises these findings and expands them with EROPA requirements from authoritative sources and a survey of DPOs to complete **RSQ1**.

3 Research Methodology and Background

This chapter provides an overview of the research methodology choices made in the thesis. Section 2.1 provides an overview of the research approach, and Section 2.2 details the research design. Section 2.3 provides an overview of the data collection and analysis approaches. Section 2.4 addresses the ethical considerations of this research, while Section 2.5 outlines its limitations. Section 2.6 provides background information and Section 2.7 summarises how the methodology addresses each research sub-question.

3.1 Research Approach

The Action Design Research (ADR) methodology [1] was selected to develop the ERoPA Approach in this thesis. The choice of the ADR methodology stems from the need for Information Systems research to respond to a dual mission: i) make theoretical contributions and ii) assist in solving practitioners' current and anticipated problems in an organisational and technical context [1], [137].

The ADR methodology is an iterative process consisting of several stages, beginning with an initial problem formulation stage. This is followed by a series of building, intervention, and evaluation (BIE) stages, with researchers, practitioners, and users each contributing to these BIE stages. For this study, the researcher leads the design while practitioners actively design, develop, and evaluate solutions related to their practice. The practitioner brings practical knowledge, experience, and understanding of the problem while learning and adapting their practice based on research findings [1]. The end user is an individual, a group, or an organisation that will directly use or benefit from the outcome, output, or results of the research [1]. The perspectives and contributions of each of these participants are gathered to contribute to the optimal design. The process concludes with a guided emergence of the optimal design, leading to a formalisation of learning [1] (see Figure 1 in Section 1.4 for a visualisation of how ADR is applied in this thesis.

ADR Roles The practitioners and user participants for each stage of the ERoPA Approach's development are selected based on their relevance and experience with that stage of the ERoPA Approach. The Practitioners are made up of the following: (i) the Data Privacy Vocabularies and Controls Community Group (DPVCG) made up of technologists and legal experts whose focus is the development of the Data Privacy Vocabulary (DPV); their expertise is leveraged to support the development of the ERoPA Approach (ii) the Upsilon Data Protection team, operating the privacy function of a multinational organisation. This contains five certified DPOs with experience across multiple organisations and industries.

The ADR Users consist of (i) data protection professionals; these are the participants who completed the survey to support the gathering of requirements for the ERoPA Approach, gathering the Data Protection Professionals' requirements. These participants hold a high level of qualifications and have a sufficient level of expertise with a blend of legal and technological experience (see Section 4.6.2) (ii) expert interview candidates; the classification of an expert must be or have been a DPO, possess a DPO qualification, and have at least five years of data protection experience (see Section 7.4.3) (iii) Industry users; commercial organisations such as Signatu [36] that actively utilises component parts developed for the ERoPA Approach. Sections 4.1.1, 5.1.1, 6.1.1, and 7.1.1 set out the specific roles for each stage of this research.

The problem formulation stage—This initial stage of the ADR process involves a literature review and organisational context research. This stage aims to identify and conceptualise the research opportunity and the problem that needs to be addressed. It identifies theoretical approaches and prior technology advances that could offer opportunities. It also identifies, secures organisational commitment, and sets up stakeholders' roles and responsibilities for research participation. This stage primarily focuses on stakeholder needs.

The Building, Intervention and Evaluation (BIE) Stage - This stage of the ADR involves researchers designing the artefact and then gathering inputs and feedback from practitioners and users to improve each iteration of the artefact development. Many BIE loop cycles can occur as the design is enhanced and optimised [1]. The learnings from each BIE stage are gathered in the ADR reflection and learning process.

The Reflection and Learning Stage – This stage draws on the principle of 'guided emergence' [1], which emphasises that the artefact will reflect not only the preliminary design created, in this case, by the researcher, but also its ongoing shaping by organisational use, different perspectives and the participants, and by the outcomes of authentic and concurrent evaluation [1]. The guided emergence involves gathering qualitative feedback from all stakeholders involved in each BIE stage of the ADR process. This information is reflected upon and documented, and practical and theoretical learnings are identified from the reflective analysis. Its information is captured as actionable insights for future processes. The empirical findings are compared to existing theories, and new theoretical contributions based on the ADR project outcomes are identified. The findings are shared with the broader community through publications, presentations, and knowledge for academic and practitioner communities to validate and refine insights.

The formalisation of the Learning stage—The final stage of the ADR process typically involves developing a systematic evaluation framework, where qualitative and quantitative research methods are analysed to identify design principles emerging from empirical findings.

3.2 Research Design

This section describes the specific design utilised for each stage of the ADR methodology. It also describes the methods chosen and the justification for their selection.

3.2.1 Design for the Problem Formulation Stage

This initial stage of the ADR process involves a literature review and organisational context research to identify and conceptualise the research opportunity and the problem that needs to be addressed. This stage identified theoretical approaches and prior technology advances that could offer opportunities. It also identified and secured organisational commitment and set up stakeholders' roles and responsibilities for research participation.

This stage primarily focused on the research sub-question **(RSQ1)** to answer the question "*What are the stakeholder requirements for the ERoPA Approach?".* It established the importance of RoPA in demonstrating compliance with the GDPR, gathering the success factors of RegTech and reviewing the State of the Art for a Semantic Web-enabled RoPA. The problem formulation stage also established the requirements that RoPA Stakeholders would have for ERoPA by synthesising data gathered from analysis of (i) regulator-supplied RoPA templates and guidance documents, (ii) commercial tools and (iii) a survey of data protection professionals.

The gathering of ERoPA requirements uses Macauley's requirement engineering approach [138] to support the gathering of ERoPA requirements. This is utilised to ensure that an iterative, systematic, efficient and effective approach to elaborating an explicit requirements specification is used. This process contains several stages, as follows: (i) Elicitation, (ii) Analysis, (iii) Specification and (iv) Validation & Verification.

A **Research Survey** was chosen to gather Data Protection Professionals' requirements of ERoPA. This survey method, based on Saunders et al. [41], is useful for exploratory research, as it enables the researcher to reach many stakeholders and gather their opinions in a structured manner. The survey uses the research onion framework process [139] to generate the question set for the study to ensure that the best practices of content, construct, and reliability are met. Ethical approval must be in place when deploying a survey.

3.2.2 Design for Building, Intervention, and Evaluation (BIE) Stages

Building on the requirements specification (see Section 4.7) for the theorised ERoPA Approach the development of ERoPA Approach is iterated over three BIE stages as follows: (i) a semantic ontology for the **representation** of GDPR Accountability concepts necessary to create and maintain RoPA (see Section 2.3.2), (ii) an enabling semantic interoperability specification for the **collection** and **transfer** of RoPA information (represented as specified by the ontology) from stakeholders such as

organisational units or data processors and relevant data protection stakeholders (see Section 2.3.3) and (iii) application requirements to enable the validity and conformance of RoPA and provide DPOs with tools for the **review** and **inspection of** RoPA in both human-readable and machine-readable formats (see Sections 3.3.3 and 3.3.4).

3.2.2.1 Design for BIE 1

The first Building, Intervention, and Evaluation (BIE) Stage of the ADR primarily addresses **RSQ2a**, "*What information is required to be maintained for the ERoPA Approach?*" and **RSQ2b,** *"How can this information be represented as a Semantic Web ontology for the machine-readable and interoperable representation of information required by the ERoPA Approach?*"

Methods for Building: The first component of the ERoPA Approach, the machine-readable representation of RoPA information, known as the Common Semantic Model of RoPA (CSM-RoPA) [13], is developed in this cycle. The terms and concepts for creating a Common Semantic Model of GDPR information required to be maintained in an ERoPA are gathered (RSQ2a). These requirements are based on state-of-the-art reviews and practice-inspired research. They are elicited from authoritative sources (legal texts, regulator templates, guidance documents, and reports), academic sources, and stakeholders' requirements to answer RSQ1 (see Section 4.2).

Building on the success factors of RegTech (See Section 2.5.1) this research uses Semantic Web technology [7], [8] for the ERoPA Approach as it enables machines to process machine-readable documents and understand web content, leading to the development of more intelligent algorithms, semantic search engines, and knowledge-based systems (see Section 3.6 for background information on the Semantic Web). To express the GDPR terms concepts found in a Semantic Web form, the **NeOn (Networked Ontologies) Methodology** for Ontology Engineering [38] was used to create the CSM-RoPA ontology.

The NeOn methodology is chosen because it is a comprehensive framework for the collaborative development of ontologies. It provides guidelines and practices for ontology development, emphasising the reuse and integration of existing ontologies and data sources. It also follows ontology engineering best practices (RSQ2b) (see Section 5.4).

There are several key benefits of the NeOn Methodology. NeOn offers efficiency by reducing the time and effort required to develop ontologies and promoting reuse and collaboration. Neon provides a flexible approach that supports various ontology development scenarios and contexts. NeOn is also suitable for creating large-scale, interconnected ontology networks, offering the opportunity to extend ERoPA beyond the domain of GDPR accountability data solely for RoPA. NeOn also enhances the quality and consistency of ontologies through systematic validation and evaluation. By following the NeOn Methodology, organisations can create robust and interoperable

ontologies that effectively support their knowledge management and Semantic Web applications. The NeOn Methodology guides different aspects of the development of an ontology. Each scenario responds to specific activities such as requirements gathering, ontology reuse, ontology engineering, design enrichment and validation. The structured approach of the NeOn methodology provides a systematic and disciplined approach to ontology engineering, ensuring the development of high-quality, reusable, and maintainable ontologies.

The **WIDOCO ontology documentation wizard** is used to document the CSM-RoPA ontology. The documentation of the ontology followed WIDOCO's use of the W3C data on the web best practices [27], [140] . The wizard was selected as it detected missing vocabulary metadata and created documentation with diagrams, human-readable descriptions of the ontology terms, and a summary of changes for previous ontology versions. The documentation consists of a set of linked, enriched HTML pages that end users can further extend. WIDOCO is open source and builds on well-established Semantic Web tools. The building of the CSM-RoPA ontology also utilises the **FAIR principles (FAIR 'Findable, Accessible, Interoperable and Reusable')** designed to meet Semantic Web standards and best practices, as it makes data discoverable, reliable, linkable and reusable across domains. [28], [141] .

Methods for Intervention: The ADR methodology fosters participatory and practical research and technology-based interventions, which use iterative cycles of planning, acting, observing, and reflecting [142]. The key methods employed in BIE 1 involve regular engagement with practitioners such as an expert group, W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), a small-scale use case, and presentations at conferences, and industry use to gather data on the effectiveness of these CSM-RoPA development [142].

The CSM-RoPA ontology was developed using three distinct cycles. Between 2021 and 2024, the mapping outcomes were presented to the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), comprising technologists and legal experts, to leverage their expertise. This process of continuous dialogue with stakeholders such as the DPVCG (Practitioners) and a practising DPO (The Researcher) offered feedback and iteratively improved the ontology to enable full RoPA concept expressivity (see Section 5.4.4).

An initial use case demonstration was conducted to show how CSM-RoPA can express a RoPA (See Section 5.4.3). A use case demonstration was selected, as it is a common approach in applied research, especially in fields like computer science, human-computer interaction, and software engineering and ensures that best practices for ontology design are met [143].

The CSM-RoPA ontology was published in four peer-reviewed publications and presented at three conferences (see Sections 1.7 and 5.4.5). Regular engagement with Signatu [36] , who

develop automated solutions for GDPR compliance, provided iterative feedback during the development cycle and now uses CSM-RoPA ontology concepts integrated in their products.

Methods for Evaluation: The CSM-RoPA model's evaluation involves assessing its technical quality and effectiveness in meeting its intended purpose, ontology validation (checking if the ontology correctly represents the real-world domain) and ontology verification (ensuring it's built according to specifications) [144]The NeOn methodology's scenario-oriented approach highlights the importance of reuse and collaboration; thus, the evaluation focuses on how effectively the ontology supports these elements and its overall suitability for particular application contexts. The evaluation consists of five key areas, each of which is discussed below:

(i) A gap analysis to establish whether the ontology meets its intended purpose and contains the necessary concepts and relationships to represent a GDPR real-world RoPA domain [145], [146].

(ii) An analysis of the extent to which CSM-RoPA ontology meets the competence questions set out in the NeOn ontology requirements specification (See Section 5.2.2) to meet user requirements [145], [146]. This evaluation approach was selected because competency questions are a well-established evaluation method in ontology engineering. This approach checks whether the ontology can adequately represent knowledge needed to answer a predefined set of questions, essentially validating the ontology's scope, correctness, and usefulness [145].

(iii) An analysis to determine whether the CSM-RoPA ontology meets Semantic Web standards and best practices based on FAIR (**'Findable, Accessible, Interoperable and Reusable'**) principles for research data [28]. FAIR principles provide guidelines and best practices for describing and accessing research data to be reused by others [28]. For the developed ontologies, the FOOPS ontology pitfall scanner [147] was used to assess whether the vocabulary (OWL or SKOS) conforms to the best practices for publishing ontologies on the Web.

(iv) This analysis using FAIR extends beyond the NeOn Methodology and provides researchers with the means to assess whether the vocabulary conforms to these best practices.

(v) A **use-case** to show how CSM-RoPA can express a RoPA. This evaluation approach was selected as it is a common approach in applied research, especially in fields like computer science, human-computer interaction, and software engineering and ensures that best practices for ontology design are met [143].

(vi) Peer Review - This approach is based on a peer-review-based approach for ontology evaluation [148]. The developed CSM-RoPA is presented to end users including legal

experts, researchers, technologists, engineers, and practitioners in the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) which met weekly throughout the ADR process. Peer feedback also occurred at two conferences and three peer-reviewed publications, where users' feedback on CSM-RoPA is gathered for the next stage of development (see Section 5.4). The use of peer review enables practitioners and users to contribute to artefact development and ensure that the evaluation findings are accurate

3.2.2.2 Design for BIE 2

Building on the CSM-RoPA ontology component of the ERoPA Approach developed in BIE1 for representing GDPR RoPA information, the second BIE loop addresses RSQ2c: *"How can the information required by the ERoPA Approach be communicated between stakeholders?"* The DPCat is developed as an enabling semantic interoperability specification for the **collection** and **transfer** of RoPA information from stakeholders such as organisational units or data processors and relevant data protection stakeholders (see Section 2.3.3) to meet the requirements for the ERoPA Approach (see Section 4.7).

The DPCat specification is conceptualised to enable the flow of GDPR accountability data between RoPA stakeholders (see Section 4.7) based on Semantic Web standards and best practices [149] and ontology engineering (see Section 3.2.2). RoPA data must be maintained in a shareable format for regulators' inspection (see Section 4.5.3), as sharing with auditors and certification bodies is increasingly important (see Section 4.7). This specification should use an agreed interoperability standard for representing, collecting, and transferring RoPA information in a machine-readable way, enabling publishers to use a standardised model and vocabulary for metadata from multiple catalogues. (see Section 4.7 req. 2.2) [13], [18], [19], [33].

Methods for Building: For the development of the DPCat specification the **NeOn Methodology (Networked Ontologies)** for Ontology Engineering [38] (which was also used in BIE 1 for the development of the CSM-RoPA ontology) is selected. The NeOn Methodology is selected as it offers a structured development approach and emphasises the reuse and reengineering of existing semantic resources, which is key to achieving semantic interoperability in the development of the specification [38] .

The DPCat specification also utilises the design pattern of the **W3C Data Catalogue Vocabulary DCAT**, and which are selected for this research. The W3C DCAT (Data Catalog Vocabulary) is a standard RDF (Resource Description Framework) vocabulary used for publishing data catalogues on the web[150]. A data catalogue is a centralised repository and metadata management tool that provides an organised and searchable inventory of an organisation's data assets. It empowers users

to discover, understand, and leverage data for analytical purposes, reporting, and informed decision-making [151], [152] (for more information on Data Catalogues, see Section 3.6).

The DCAT design pattern was selected for the DPCat specification as it facilitates the interoperability of data catalogues, enabling datasets to be discovered and accessed across different platforms and domains . It helps describe datasets, data services, and catalogues in a structured, machine-readable format, making it easier for data portals, repositories, and other platforms to interoperate. The key features of DCAT [153]:

- Interoperability: DCAT ensures that data catalogues from different organisations or domains can be described using a common format, improving discoverability and access.
- Metadata Standardisation: It provides a standardised way to represent metadata about datasets, such as their title, description, publisher, and distribution format.
- Dataset Discovery: By making datasets easily discoverable through standardised descriptions, DCAT enables data consumers to find relevant datasets across various catalogues.
- Linked Data Friendly: As part of the RDF family, DCAT supports linking datasets to other web resources, making it suitable for the Semantic Web and Linked Data applications.

DCAT-AP stands for DCAT Application Profile for data portals in Europe. It is a specialised profile of the DCAT specification created by the European Commission to ensure interoperability between European open data portals [154]. DCAT-AP provides a set of rules, guidelines, and extensions on top of the W3C DCAT vocabulary, with its main goal to make it possible to federate datasets from different portals into the European Data Portal. DPCat reuses the design approach of DCAT-AP to further constrain the DCAT vocabulary by defining a compliance profile of Mandatory, Recommended and Optional fields and using SHACL (Shapes Constraint Language) tests to operationalise those tests.

Methods for Intervention: The key methods employed in BIE2 involve regular engagement with practitioners such as the DPVCG expert group, a case study, journal publications, and industry use to gather data on the effectiveness of these DPCat development [142]. The DPCat Specification was developed in conjunction with a practising DPO (the Researcher) and the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), comprising technologists and legal experts (Practitioners). Similar to BIE1, this continuous dialogue with these stakeholders offered feedback and iteratively improved the specification.

A case study implementation was conducted that examines five common GDPR scenarios for gathering and transferring GDPR RoPA compliance information among stakeholders (U1-U5), based on the RoPA of the European Data Protection Supervisor (EDPS) [18]. These standard

transfers and use cases are based on real-world RoPA information representation and communication of RoPA information among stakeholders, gathered from the Data Protection Professionals survey (see Section 4.6.2). A case study was selected as it is an empirical inquiry, which examines a contemporary phenomenon in its real-life context [155], which allows researchers to delve into the essential characteristics, meanings, and implications of the case.

Methods for Evaluation: The evaluation of the DPCat interoperability specification component of the ERoPA Approach determined the extent to which DPCat meets the Interoperability specification competence questions. This evaluation approach involves a case study involving five common GDPR scenarios to gather and transfer GDPR RoPA compliance information among stakeholders based on the established use cases (U1-U5).

To demonstrate the capability of DPCat to meet these competencies and use cases, a series of RoPA information representation and gathering tasks taken from the Data Protection Professionals survey (see Section 4.6.2), are conducted using DPCat in the case study. The gathered RoPA information is validated using Shapes Constraint Language (SHACL) to ensure the correctness of information as per DPCAT-AP specifications and DPCat requirements to ensure data correctness. In this use case, the open-source and freely available TopBraid [156] SHACL tool is utilised for executing the constraints

Similar to BIE 1, a peer-review-based approach [148] for the DPCat deployment is carried out. The developed DPCat is presented to end users such as legal experts, researchers, technologists, engineers, and practitioners and the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) and is published in peer-reviewed journal publications and a conference proceedings publication (see Section 6.7.1). The learnings are gathered in the second BIE stage's reflection and learning ADR process for the next stage of development (see Section 6.6).

3.2.2.3 Design for BIE 3

Building on the knowledge gained in the first and second BIE stages regarding the development of ERoPA, the third BIE stage addresses **RSQ3** "*To what extent does the ERoPA Approach support implementing GDPR accountability?* " and **RSQ4 "***What are the key considerations for organisations implementing an ERoPA Approach*? ", to identify the tools and methods and the capabilities granted by the ERoPA Approach.

Methods for Building: A deployment of ERoPA in an organisational context was conducted in BIE 3. The unit of analysis for the case study consists of the business units and partners of the Upsilon company involved in ROPA-related data protection activities. A case study was selected as it is an empirical inquiry, which examines a contemporary phenomenon in its real-life context [155]. It

allows researchers to delve into the essential characteristics, meanings, and implications of the case. Yin's case study methodology was selected as it provides a structured and rigorous approach to investigate phenomena in their real-world context [155]. A case study is well-suited for action design research (ADR) because it facilitates the exploration of a phenomenon in its practical setting. This enables researchers to iteratively develop and enhance design interventions while gaining insights from the process. This methodology aligns with ADR's objectives of creating and refining solutions while also generating knowledge regarding the design process itself.

The building steps required for the Case Study deployment (see Section 7.5.1) brought together five components of the ERoPA Approach, which are listed below. Together, these components make up the ERoPA Approach to support the case study deployment.

A Data Catalogue was required to organise, store and manage datasets. Data catalogues support metadata management for describing datasets effectively. They offer robust search facilities and facilitate the publishing of data sets. (For more information on data catalogues, refer to Chapter 2.6). The catalogue in the **DPCat** specification utilises the DCAT and DCAT-AP pattern to express RoPA information in a predefined structure for exchanging information (see Section 3.2.2.2). When data is presented in the DPCat catalogue structure, it benefits from a data catalogue providing the ability to publish and query datasets using specialised tools in a triple store.

A tool for Data Conversion to RDF was required to create a schema to support the conversion of data presented as a spreadsheet into RDF format. This was achieved using the **OntoRefine** tool [157] to create a schema and apply the schema to a spreadsheet to convert the data into RDF format suitable for loading to the knowledge graph. OntoRefine was selected for the data transformation as it is widely used in semantic data projects, particularly for preparing data with knowledge graphs and other linked data systems. OntoRefine provides mapping rules for semantic ontology alignment to support RDF (Resource Description Framework) transformations.

A Triple Store was required to store RDF triple data. The **Graph DB** [158] tool was selected as it provides built-in tools to assist in importing RDF into the knowledge graph triple store. These imports can come as a URL link, an upload file, or a clipboard text string directly entered into the upload tool. This gives the user an easy-to-use interface for loading RDF data to the knowledge graph. Graph DB also provides SPARQL query as an inbuilt feature.

A Data Quality Assurance Specification was required for conformance and provenance checking to ensure that input to RoPA provided by stakeholders met data quality rules. Similar to BIE2 the **SHACL** data shapes ontology was used to create preferred graph shapes, which were used to identify data quality issues and non-compliances

A Query Tool was required to support the DPO in conducting typical data protection tasks. The query engine SPARQL was used for these processes (similar to BIE2). A set of standard queries based on typical DPO tasks was used to conduct four typical DPO tasks (see Section 7.5.1).

Methods for Intervention: A series of use cases were performed on the case study organisation prototype, and observations were gathered through the deployment. ERoPA was used to conduct typical DPO tasks and to support a regulator accountability framework.

Gathering observations during the case study deployment is crucial for gaining in-depth insights into real-life contexts. The process used for gathering observational feedback through the deployment following Morgan's framework [29]. Such observational data can provide rich, qualitative insights that support or enhance other data types, such as interviews or document analysis. Observations in case studies allow for an authentic view of the context and behaviours, adding a depth of insight beyond what interviews or surveys alone can provide. By carefully planning and structuring observations, one can ensure relevant and reliable data collection that enriches the overall case study analysis [29]. Analyse and synthesise observations regularly to identify patterns or themes for subsequent observations.

The case study deployment was presented to data protection experts in semi-structured interviews. This method was chosen as semi-structured interviews provide an effective balance of flexibility and focus, making them ideal for collecting in-depth information, particularly when personal insights are essential [159]. This method blends predefined questions with the opportunity to delve into topics based on participant responses, allowing interviewers to probe deeper into noteworthy points and seek clarification [160]. This interview format encouraged participants to articulate their thoughts in their own words, resulting in richer data that might be overlooked with closed questions. The conversational style fosters a comfortable environment that promotes openness. Core questions facilitate easier comparisons across participants, supporting efficient analysis. The interviewer gathered notes from these interviews, and they were analysed using pattern grouping to identify common themes [161]. This method organises and makes sense of data by identifying patterns and relationships. It was chosen because it is valuable for analysing qualitative data due to its flexibility and helps in summarising, interpreting, and understanding datasets by revealing underlying structures and insights.

Methods for Evaluation: The evaluation of the ERoPA case study deployment utilised a case study to provide an empirical inquiry in a real-life context [155]. The key benefits of a case study are outlined at the start of this section. The case study consisted of three evidence bases collected and synthesised to gather key findings. The first data collection was direct observations gathered during the case study deployment, following Morgan's framework [29] (see Section 7.4.2). These observations and synthesised common patterns were identified. The second data collection consisted of semi-structured interviews with data protection experts. Semi-structured interviews

were selected for their effective balance of flexibility and focus, making them ideal for collecting in-depth information, particularly when personal insights are essential [159]. The third data collection in the case study validated a real-world use-case for its extent to satisfy the UK GDPR regulator (ICO) Accountability requirements. The analysis of the case study findings provided key learning for the development of the ERoPA Approach (see Section 7.7).

The **ICO Accountability Framework** is utilised to evaluate the extent to which the ERoPA Approach supports implementing GDPR accountability (RSQ3). The Information Commissioner's Office (ICO) is a United Kingdom Government body that regulates the UK GDPR. The ICO launched the Accountability Framework in 2020 to help organisations assess their compliance with key requirements under the GDPR [68]. The ICO accountability framework has several uses for organisations, such as recording, tracking, and reporting compliance progress [162]. It can check the organisation's existing practices against the ICO's expectations to identify areas for improvement and clearly understand how to demonstrate compliance and increase senior management engagement and privacy awareness. The ICO accountability framework is chosen for this research, as it is a method of evaluating an organisation's GDPR compliance and demonstrating GDPR accountability (for more information on the ICO Accountability Framework, see Section 3.6).

The observations, interview findings and ICO Accountability Framework of the case study deployment were analysed to evaluate the extent to which the ERoPA Approach supports implementing GDPR accountability (RSQ3).

3.2.3 Design for the Formalisation of the Learning Stage.

The final stage of the ADR process addresses **RSQ4**, "*What are the key considerations for organisations implementing an ERoPA Approach? ".* The generalised research outcomes are presented as a Zachman Framework [30], providing guidelines for organisations considering ERoPA deployment.

The Zachman Framework is selected as it provides a comprehensive conceptual enterprise model [30]. The findings from the (i) State of the Art review (ii) Data Protection Professionals survey (iii) CSM-RoPA development (iv) DPCat use case (v) ERoPA deployment observations (vi) expert interviews, and (vii) accountability verification exercise are brought together and presented in a Zachman framework, based on 2001 version of the framework [163] (see Section 7.8.3). This Zachman framework version provides a structured approach to understanding and defining the various components involved in development, allowing teams to effectively transform this framework into practical and useful applications. By mapping out different perspectives and aspects of a project, the framework helps ensure that all elements are considered, leading to more coherent and successful outcomes [1], [30]. The Zachman Framework is a systematic method for

comprehending and recording an organisation's architecture by addressing six essential aspects (What, How, Where, Who, When, and Why) and six perspective rows to represent stakeholders' viewpoints. Whilst there are different types and variations of the Zachman Framework, the version of the Zachman Framework used in this thesis is based on the standard Zachman Framework (2001). This version uses the stakeholder dimensions of Planner, Business Owner, Developer, Builder and User (DPO)[163] .

In the version of the Zachman Framework used in this Thesis, the builder and technician roles are amalgamated into one role, as these roles are broadly aligned in the development process of the ERoPA Approach. The key benefit of the Zachman Framework is that it provides a holistic perspective on the whole enterprise while allowing focus on specific aspects of the object being built. The Zachman approach offers many benefits, including efficient resource allocation, improved decision-making, better risk management, reduced maintenance costs, increased agility, and better IT alignment with business objectives.

3.3 Data Collection and Analysis Methods

This section describes the data collection and analysis methods and justifies the selection of each method.

Desk Research - This is a widely used data collection method in this thesis [164]. The researcher has examined the state-of-the-art literature to gather relevant data, such as RoPA templates and guidance documents provided by Regulators. This data collection method was selected as it allows for a broad, comprehensive overview of the research topic.

Survey Data - This research method is utilised to gather quantitative and qualitative data during the problem formulation phase to ensure that the issues faced by Data Protection Professionals. Surveys offer a structured approach to gathering information from key users to inform this research [139].

Correspondence Patterns Representation - Using Scharffe's mapping classifications, this method supports the analysis of matching GDPR concepts to Data Privacy Vocabulary (DPV) terms. This exercise aims to identify exact, narrower, or no matches. This process drew on François Scharffe's 2009 work on ontology alignment [24].

Case Study Observations - this research utilises Morgan's framework [29] to gather observations through the Upsilon case study (see Section 7.4.2). The process was selected as it provides a structured approach, from gathering observational feedback through deployment to synthesising observations to identify common patterns within the observed data.

Expert interview Data - This research gathers opinions using semi-structured interviews [159] with data protection experts. The decision to use semi-structured interviews would benefit from a smaller, more experienced group of professionals, thus providing greater insight when discussing the implementation of ERoPA. The open nature of semi-structured interviews allowed for additional exploration of responses where appropriate. This allowed the interviews to develop beyond the structured question to explore themes relevant to ERoPA deployment. The interview notes were recorded and stored on a secure Google Drive.

Pattern Analysis - This method was utilised to organise and make sense of data gathered in the expert semi-structured interviews by identifying patterns and relationships. It was chosen because it is valuable for analysing qualitative data due to its flexibility, and helps in summarising, interpreting, and understanding datasets by revealing underlying structures and insights [161].

3.4 Ethical Considerations:

Ethical issues in this research were addressed as follows : (i) Were participants fully informed about the purpose, risks and benefits of this research? This was addressed in the survey of Data protection professionals and the semi-structured interview held with experts, firstly by providing a transparency notice with clear, concise, and comprehensive information about the research, including its purpose, procedures, risks, and benefits, before they consented to participate and secondly, both processes were reviewed and approved by the Dublin City University Ethics Committee (see Appendix A), (ii) Was the privacy and confidentiality of participants protected so that there could be no identification or re-identification of participants? This was addressed by fully anonymising the information gathered and securely storing the gathered data.

3.5 Limitations

The study design has four methodological limitations that the reader should carefully consider in interpreting the findings. These limitations have been mitigated where possible to reduce any impact on the research, and are discussed below:

1. The translation of the 12 non-English language regulator RoPA templates was completed using Google Translate [165]. Whilst the output of these translations has been published in peer-reviewed publications [13], [18], [19] further refinement of the information gathering process may be possible via native speaker evaluation of non-English language regulator templates.
2. The survey of Data protection Professionals utilised DPO peer networks and LinkedIn posts to gather participants from industry, consultancy, and academia.

While the survey participants were suitably qualified participants the network peer groups were largely based in Ireland and thus may not represent wider geographic opinion.

3. The Upsilon case study was conducted for one organisation only, albeit that Upsilon is a large, complex organisation; therefore, the results may not be representative of other organisations and might not be transferable to different situations.
4. The expert group that completed the semi-structured interviews consisted of five experts with significant experience. Still, the size of the sample group may be considered small, and it could be extended in future work.

Organisations considering ERoPA deployment should take into account such limitations, as the ERoPA Approach has not yet been proven in commercial live deployment.

3.6 Background Information

This section provides additional background information on technologies and techniques used in the artefact creation and evaluation process.

The Semantic Web: The Semantic Web is an extension of the World Wide Web, envisioned by Tim Berners-Lee, the inventor of the World Wide Web [5]. The Semantic Web aims to make web content machine-readable and interpretable by computers, enabling them to process and interpret information more easily and effectively. In the context of the Semantic Web, the word 'semantic' indicates machine-processable [5], [8]. Semantic Web uses a structured data modelled as a graph and publishes it in a way that allows interlinking across servers. The Semantic Web relies on standardised technologies and languages, such as RDF (Resource Description Framework)[149] , RDF Schema (RDFS) [166], and OWL (Web Ontology Language) [25] , to represent data and metadata in a structured and semantically rich way. These technologies allow building machine-readable ontologies to create relationships and properties among concepts and entities. Semantic interoperability ensures that the precise format and meaning of exchanged data and information are preserved and understood, enabling better integration across different datasets [109], [167]. Some of the key components and concepts of the Semantic Web include [8] :

URI (Universal Resource Identifiers) - URI provides a global naming convention (that drives the Web's benefits. URIs have an international scope and are interpreted consistently across contexts [8]. Associating a URI with a resource means anyone can link to it, refer to it, or retrieve a representation of it. URIs use namespace prefixes to group related identifiers in an ontology. This simplifies referencing of URIs by providing a shorthand notation. An example of a namespace would

be '@prefix ex: http://example.org/ontology#,' where ex is used as the shorthand for the namespace, making it easier to refer to resources.

RDF (Resource Description Framework) - RDF is a standard model for representing data and metadata on the web[149], [166] . It uses a triple structure (subject-predicate-object) to describe relationships between resources. Subjects, predicates, and objects (except data values) are referred to by URIs in RDF. An example of a sentence with a subject, predicate and object would be 'the cat crossed the road', where 'the cat' is the *subject* who performs the action, 'crossed' is the *predicate* being the action or verb, and the 'road' is the *object*, being the thing that received the action. RDFS is a specification enabling the definition of RDF vocabularies. The semantic knowledge represented in RDFS can be stored in triple stores and queried through SPARQL, RDF's query language (and protocol) [166] .

Triple Stores - An RDF triple store is a graph database that organises data as a web of objects and employs inference to reveal new insights from existing relationships [168] . Its flexible and dynamic characteristics enable the connection of various data types, indexing them for semantic searches and enhancing them through text analysis to create large knowledge graphs.

OWL - The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies[25] .

SKOS - The Simple Knowledge Organization System (SKOS) is a W3C recommendation for representing thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary [169] .

Ontologies - Ontologies define vocabularies and relationships among concepts in a specific domain. They provide a formal and structured representation of knowledge, enabling machines to interpret the meaning of data and reason about it.

Linked Data - Linked Data is a set of best practices for publishing and interlinking structured data on the web. It encourages using standard formats and protocols to create a global network of interconnected data [170].

SPARQL (SPARQL Protocol and RDF Query Language) - SPARQL is a query language for querying RDF data [171] . It allows users to retrieve and manipulate data stored in RDF format using a syntax like SQL (structured query language).

SHACL (Shapes Constraint Language) - SHACL is a World Wide Web Consortium (W3C) standard language for describing RDF graphs [112]. SHACL has been designed to enhance the semantic and technical interoperability layers of ontologies expressed as RDF graphs. As with RDFS schemas and OWL ontologies, SHACL contains metadata about datasets, but this metadata serves a different purpose: to help validate the data instead of enabling inferencing. SHACL describes a dataset's structure by listing classes, properties, and relationships that a dataset must conform to. SHACL

shapes are developed and used to represent the cardinality and type constraints to ensure correctness with defined requirements.

Inference and Reasoning - Semantic Web technologies enable machines to perform inference and reasoning based on data semantics and ontologies to derive new facts. This allows for the automated deduction of new knowledge from existing data.

The Semantic Web was selected for the ERoPA Approach as it offers improved data integration, enhanced search capabilities, and more intelligent applications [8]. It enables machines to process and understand web content more sophisticatedly, leading to the development of more intelligent algorithms, semantic search engines, and knowledge-based systems [172].

Data Catalogues: A data catalogue is a centralised repository and metadata management tool that provides an organised and searchable inventory of an organisation's data assets. It empowers users to discover, understand, and leverage data for analytical purposes, reporting, and informed decision-making [151], [152]. Acting as a bridge between data producers—such as data engineers and scientists—and consumers, including analysts and business users, data catalogues enhance data governance and utilisation. Some key advantages of data catalogues [151], [173] include:

- The capability to store essential metadata, which encompasses information about data sources, business definitions, relationships, schemas, lineage (tracking the origin and processing of data), and quality metrics.
- User-friendly interfaces that facilitate search and discovery.
- Assistance in enforcing governance policies by documenting ownership, access permissions, and compliance requirements for datasets.
- Transparency in the flow of data through systems, from source to final form, fostering trust and auditability.
- Seamless integration with other tools and platforms, including data warehouses, business intelligence (BI) tools, and cloud environments, which enhances accessibility and usability.

DCAT-AP (DCAT Application Profile) is the EU's effort to standardise catalogue metadata in data portals [174]. DCAT-AP describes public sector datasets in the EU's Open Data portals. It enables cross-data portal search by harmonising the metadata collected and enabling common metadata collection and search for diverse datasets. This is achieved by exchanging standard descriptions of datasets among data portals. In addition, DCAT-AP proposes mandatory, recommended, or optional classes and properties for a particular application. Some examples of this for a DCAT-AP dataset would be that 'description' is mandatory, 'contact point' is recommended and 'access rights' is optional [154].

The ICO Accountability Framework utilised in the Upsilon case study deployment (see Chapter 7) contains ten specific GDPR accountability categories [162]. Each category of the ICO framework includes several expectations (of how an organisation can demonstrate accountability), and each of the 77 expectations contains many detailed questions (see Table 9) [162].

Table 9 Overview of the ICO Accountability Framework

ICO Category	No. of Expectations	No. of Questions
Leadership and Oversight	6	33
Policies and procedures	4	17
Training and Awareness	5	17
Individuals' rights	11	42
Transparency	7	31
Records of processing and the lawful basis	10	33
Contracts and data sharing	9	31
Risks and Data Protection Impact Assessments.	5	29
Records management and security	12	63
Breach response and monitoring	8	38
Totals	77	334

An example of an expectation is expectation 6.2.2: 'Your organisation regularly reviews the record against processing activities, policies and procedures to make sure that it remains accurate and up-to-date, and you clearly assign responsibilities for doing this.' The framework provides the necessary detailed granularity that enables an organisation to evaluate their level of compliance relative to each statement using a four-level scale, which ranges from not meeting/ partially/ fully meeting this expectation or as 'not applicable.'

3.7 Chapter Summary

This chapter presents an overview of how the ADR methodology supports the iterative development of the ERoPA Approach. The problem formulation stage of ADR gathers information from desk research, a review of organisational approaches, and a survey of Data Protection professionals to identify the stakeholders' requirements for the ERoPA Approach (**RSQ1) and** what information must be maintained in a RoPA (**RSQ2a**).

Following the identification of the requirements for the ERoPA Approach, there are three **building, intervention, and evaluation (BIE)** stage loops in this thesis. Each of these ADR loops supports the iterative development of the ERoPA Approach. The first of these BIE loops identifies what information must be maintained in a RoPA (**RSQ2a**) and what steps are required to develop an ontology to represent ERoPA information (**RSQ2b).** The second BIE loop builds on the ERoPA ontology developed in BIE 1 to identify the steps required to create a specification for the interoperable exchange of ERoPA information between stakeholders (**RSQ2c).** This is necessary to support the collection, transfer, and review of the ERoPA Approach. The third BIE further develops

the ERoPA Approach with a case study deployment of ERoPA to establish the extent to which the ERoPA Approach supports implementing GDPR accountability (**RSQ3).**

The final stage of the ADR **methodology** is referred to as **the Reflection and Learning Stage,** which identifies the key considerations for organisations implementing an ERoPA Approach (**RSQ4).** This is achieved by synthesising the findings from the thesis together and presenting them in **a** Zachman framework for the ERoPA Approach, offering guidelines for organisations considering deployment to demonstrate accountability.

4 Requirements for the ERoPA Approach

4.1 Chapter Overview

This chapter builds on the problem formulation stage of the ADR methodology (see Chapter 3) to conceptualise the ERoPA Approach as a solution to the practical problem of RoPA maintenance faced by organisations. The chapter introduces the ERoPA Approach artefact and identifies the specific stakeholder requirements that the designed ERoPA artefact will address.

4.1.1 ADR Roles for this Chapter

The researcher leads the design while organisational participants contribute their expertise and context to the development and implementation of the artefact. These roles are set out in Table 10, where the Researcher, a practising DPO with experience in conducting the planning and implementation of the ERoPA, conceptualises the ERoPA. The Practitioner role is met by the Data Privacy Vocabularies and Controls Community Group (DPVCG)[17], who contribute their legal and technical expertise and context to the development and implementation of the artefact. The user role is met with a survey of Data Protection Professionals.

Table 10 ADR Roles for the Problem Formulation Stage

Role	Assignment
Researcher	The Thesis Researcher (a practising DPO)
Practitioner	Data Privacy Vocabularies and Controls Community Group (DPVCG)
User	Data Protection Professionals

This chapter also utilises a requirement engineering approach [138] to identify requirements for the ERoPA Approach gathered from six sources. It addresses RSQ1 to determine the stakeholder requirements for the ERoPA Approach. This chapter also addresses RSQ2a to identify the information required to be maintained in an ERoPA. This chapter is based on four of the researcher's existing publications [13], [18], [33], [34].

[17] https://www.w3.org/groups/cg/dpvcg/

4.2 Requirements Gathering Process for the ERoPA Approach

The ERoPA Approach is based on utilising existing successful approaches from RegTech (see Section 2.5.1) to create an artefact that addresses the practical maintenance challenges faced by organisations regarding GDPR Records of Processing Activities(RoPA). It builds on four RegTech success factors, which are (i) the digital transformation of data, (ii) the adoption of new technologies (iii) an agreement on Common Standards and semantics and (iv) an enabling regulator (see Chapter 2.6).

The ERoPA Approach supports five key stakeholder activities: (i) the **collection** of RoPA information from stakeholders such as organisational units or data processors, which involves the gathering of RoPA information from all sources (such as Data Processing Agreements and Privacy Notices and, Policies and Procedures for example) (see Section 2.3.3). (ii) the **representation** of this information (see Section 2.3.2), in a machine-readable format (iii) the **transfer** of the information between stakeholders whilst complying with provenance and conformance rules (see Section 2.3.3) (iv) the **review** of this information using tools (see Section 2.3.4) and (v) the **inspection** of this information in both human readable and machine- readable formats (see Sections 2.3.3 and 2.3.4).

To achieve GDPR RoPA information collection, representation, transfer, review, and inspection, the ERoPA Approach requires identifying key methods and tools, referred to as 'components', that must be developed. Based on the State-of-the-Art findings, and the opportunities offered from RegTech (see Section 2.5.1), the researcher conceptualises the requirement for a semantic ontology to represent RoPA information and a specification to enable the interoperability of RoPA Information (see Section 2.6). These components are developed explicitly for the ERoPA Approach. The researcher conceptualises other tools and methods that could be used to support the validity and conformance of RoPA and provide DPOs with tools for the review and inspection of RoPA. The researcher also contemplates a set of deployment guidelines to support organisations considering the deployment of the ERoPA Approach.

As a primary step in developing the ERoPA artefact, a set of requirements for the ERoPA Approach is gathered. This section describes the process used to formulate these requirements. A requirement is defined as a constraint, ability or characteristic that a stakeholder requires for a product or process to solve a problem or reach a goal [175]. A system must satisfy a contract, a standard, a specification, or other specified formal documents [138].

This thesis uses a requirement engineering approach [138] to ensure that an iterative, systematic, efficient and effective approach to elaborating an explicit requirements specification is used. This process contains several stages, as follows: (i) Elicitation (ii) Analysis (iii) Specification and (iv) Validation & Verification.

In this research, three data sources, (i) authoritative, (ii) academic, and (iii) practice, are used to elicit data requirements. These sources are listed below.

- Authoritative - requirements gathered from legal texts, regulator templates, guidance documents, and reports.
- Academic - requirements gathered from academic research in RoPA and GDPR compliance automation.
- Practice: requirements gathered from industry commercial tools or initiatives and a survey of DPOs (see Section 4.6).

The three data sources used in this research provide a comprehensive coverage of the RoPA requirements and use, providing a solid basis for gathering requirements. Primary research methods are used to elicit the data requirements from these sub-sources. These methods are as follows:

- Conceptual research identifies and analyses the fundamental concepts underlying a problem or issue. This approach explores ideas, theories, and hypotheses and explains the relationships between concepts and their implications.
- Desk research involves gathering information by consulting documentary sources such as books, scientific articles, reports, websites, social media, etc. It is essential for understanding a topic and obtaining accurate and up-to-date information.
- Quantitative research studies data and statistics to determine trends and relationships between variables. It uses statistical tests, surveys, questionnaires (open-ended and closed-ended), and mathematical models to analyse data on a sample of individuals.
- Qualitative research is a research methodology that focuses on understanding individuals' opinions and attitudes. It examines people's thoughts, feelings, and motivations. Techniques such as interviews, focus groups, observations, and document analysis can be used. The qualitative approach is used to develop deeper insights into a subject.

4.3 Requirements Elicitation Techniques

In this research, three data sources are used to elicit data requirements for the semantic model of RoPA (see Section 4.2). The sources will identify the information required to create and maintain the RoPA document. The three categories of data sources contain six specific data sub-sources (see Table 11) used in this research. This breadth of data sources provides a comprehensive coverage of the automation of RoPA for GDPR accountability, providing a solid basis for gathering requirements. The Research methods are used to elicit the data requirements from these sub-sources. These methods include Conceptual research, Desk Research, Qualitative research studies, and Qualitative

research. Please refer to Table 11 for data sources/ sub-sources and the analysis techniques employed.

4.4 Gathering the Requirements for Machine-Readable RoPA

The research gathers the requirements for ERoPA for GDPR compliance automation from six sources, as set out in Table 11, to understand the criteria for ERoPA.

Table 11 Sources used to Gather the Requirements for ERoPA

Source type	**Research Source**	**Analysis Technique used**
Authoritative	RoPA requirements as set out in legal texts, i.e. GDPR Article 30	Desk
Authoritative	Minimum RoPA requirements determined from Regulator RoPA templates	Desk/ Conceptual
Authoritative	Findings from Regulator RoPA inspection reports	Desk
Academic	Academic research in the Area of RoPA and GDPR compliance automation	Desk
Practice	Industry / commercial tools and initiatives	Desk
Practice	Survey of Data Protection Professionals	Qualitative

Each of these sources will be analysed in the next section to determine the specific ERoPA requirements for RSQ1.

4.5 Analysis of the Authoritative Sources for Requirements for ERoPA

This section will analyse the authoritative sources of legal texts, regulator templates, guidance documents, and reports to elicit the requirements for an ERoPA.

4.5.1 Legal Texts

Legal texts (Art.30) are one of the authoritative sources from which requirements for ERoPA are elicited. The ERoPA must be capable of expressing all the concepts found within the authoritative section of the legal text of the GDPR Article 30 concerning RoPA. Please refer to section 3.3 for an overview of RoPA and Article 30 text. For this purpose, an analysis of the concepts relevant to the GDPR Article is conducted. Each clause in GDPR Article 30 is examined to identify the legal text's

relevant GDPR information categories (concepts). To derive these, the researcher performs term extraction, semantic analysis, term frequency enumeration, de-duplication, and antonym/homonym identification [24]. The criteria for inclusion or exclusion of terms are based on whether the characteristics of the term are defined within the GDPR text. If the term is not defined within the GDPR text, it is excluded from the list of information categories. The outcome of this analysis is presented in Table 12. This work was published in peer-reviewed publications and presented at conferences. The work has also been given to the Data Privacy Vocabulary Community Group W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) expert group[18]. Further, it is supported by several publications by the researcher [13], [33].

Table 12 An Analysis of Concepts Found in GDPR Art. 30 Text

GDPR Section	**Article 30 requirement**	**Concepts**	**Mandatory**
30-1a	the name and contact details of the controller and, where applicable, the joint controller, the controller's representative, and the data protection officer;	Data Controller, Representative. Joint Controller, Data Protection Officer, Name and Contact details of the Controller	Y
30-1b	the purposes of the processing;	Purpose, Personal Data Processing	Y
30-1c	a description of the categories of data subjects and the categories of personal data;	Data Subject, Personal data categories,	Y
30-1d	The categories of recipients to whom the personal data have been or will be disclosed, including recipients in third countries or international organisations.	Recipients, Third Countries, International organisations	y
30-1e	where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers, the documentation of suitable safeguards[19];	Third Countries, International organisations, safeguards, Transfer	N
30-1f	where possible, the envisaged time limits for erasure of the different categories of data;	Time limits for erasure	Where Possible
30-1g	Where possible, a general description of the technical and organisational security measures[20].	Technical and Organisational security measures	Where Possible

[18] Data Privacy Vocabularies and Controls Community Group (w3.org)
[19] Article 49(1)
[20] Article 32(1)[90]

The analysis of the legal text requires that ERoPA be capable of expressing all the concepts found in GDPR Article 30. This is the first of the requirements for ERoPA (see Section 4.7 Requirement 1.1).

4.5.2 Regulator RoPA Templates

The previous section identified the GDPR concepts that ERoPA must contain based on an analysis of the GDPR Art 30. In this section, a review of RoPA templates provided by data protection regulators is conducted to identify any additional GDPR concepts that may be requested by regulators to be documented on RoPA. A review of the State of the Art showed that no -one else has completed a similar exercise, which makes this a novel and unique contribution.

This analysis aims to identify the additional fields related to what the DPAs considered best practices to assist organisations in collecting and representing information from their various business processes. In this section, RoPA templates from all EU DPAs are examined, and a list is made of requirements that ERoPA must represent.

A detailed analysis of all available regulator-supplied RoPA templates has been conducted to complete the research in late 2021. The GDPR has 31 DPAs representing nations and member states from the EU and the EFTA EEA[21]). Each DPA provides guidance documents for RoPA based on GDPR Art.30, while some DPAs also offer templates to assist organisations with maintaining their RoPA documents. An analysis of the 31 DPA websites found that 17 DPAs provided RoPA templates, while 14 DPAs did not provide a template.

The seventeen templates identified vary in language and content. The researcher determined that five of the seventeen templates were in English. The researcher utilised Google Translate to convert the 12 non-English language templates to English [165]. While this approach to translations has limitations, the output of these translations have been peer-reviewed and presented in several publications [13], [18], [19], [33].

[21] (Note: based on EDPB membership, this consists of 32 sources. These are DPAs from 27 EU member states, plus the United Kingdom, the EDPS, and three additional members comprising the EFTA EEA states; the German regional DPAs were considered part of the national German DPA

Table 13 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	×	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×
30.1	Processing Status	×	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Tech/Org measures implementation	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Security measures	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
32	Technologies used	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
33.5	Data Breach	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

GDPR	Field	A.30	BE	GR	GB	PL	CY	FR	PT	DE	DK	LU	FI	CZ	HR	IT	LT	SI	SK
35	Risk assessment and mitigation	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
35	Relevant DPIA	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
35	DPIA Results	×	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×
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	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

GDPR	Field	A.30	BE	GR	GB	PL	CY	FR	PT	DE	DK	LU	FI	CZ	HR	IT	LT	SI	SK
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

The analysis of the DPA templates shows that the DPA RoPA templates all contain the mandatory Article 30 requirements. Table 13 presents the analysis of RoPA requirements identified from GDPR and DPA Templates. The table compares information fields in the 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 3166-2 Alpha-2 codes[22].

A review of the templates shows notable differences between them. Whilst all templates require the minimum Art. 30. Information (see table 13, column A.30), all templates go beyond this minimum requirement. An example of this would be ' legal basis', which is not a requirement of Art. 30 but is a requirement on all templates. The templates require more information than the mandatory legal requirement in all cases. The scale of additional fields can be seen in Figure 8, which shows that the number of fields required differs from 17 fields to 32 fields on some ropa templates. An example of this is as follows : Art.30 minimum 16 fields, Slovakia template 17 fields, Belgium template 32 fields. Whilst all templates contain the required Art. 30 (16 fields) requirements, the templates tend to be inconsistent (see Figure 8 for an overview of the number of fields per template, which shows the extent to which regulator templates differ from each other in terms of the number of fields of data required). This represents a challenge in producing a 'collective understanding' of what information is required to maintain a RoPA. The additional data requirements for RoPA templates are explained by regulators such as CNIL who advise to complete the additional fields in order to make 'a more global complying tool' [48]. Similarly, the DPC advises that RoPA records should contain information beyond Art. 30 such as legal basis, breach information, transfer mechanism for transfers and risk ratings for each processing activity [11].

[22] The purpose of ISO 3166 is to define internationally recognized codes of letters and/or numbers that can be used when one refers to countries and their subdivisions.

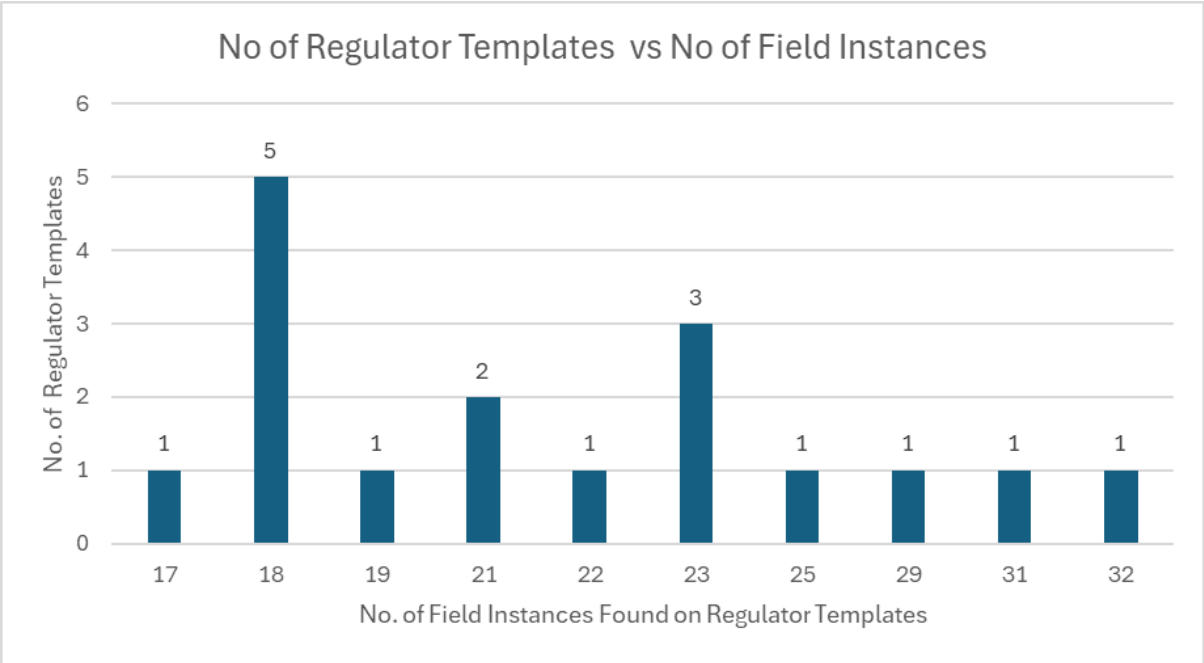

Figure 8 Overview of Number of Regulator Templates vs. No of Fields Instances Found

The researcher performed the same process as with the legal text (see 5.5.1) to identify concepts (see Fig 8 for an overview of the number of regulator templates vs. the number of field instances found). The research identified that there is a lack of consistency across regulator templates. A key finding is that all regulator templates require more than the minimum requirements of Article 30. Six regulator templates restrict their contents to conform only to the requirements specified in GDPR Art.30, with one additional field representing the 'legal basis' (see Art. 6.1). These regulator templates are from the Czech Republic, Hungary, Italy, Lithuania, Slovenia, and Slovakia. The analysis of regulator templates finds that the number of fields in a template can extend from seventeen fields, such as from Slovakia's DPA to more extensive templates, such as those provided by Belgian DPA (32 field instances) and Greek DPA (31 field Instances). The analysis of all fields across analysed templates yielded forty-seven unique concepts representing information to be recorded in a RoPA. These concepts were identified using GDPR as a primary domain, and have been validated in multiple publications [13], [18], [19], [33] and by the expert group of DPVCG. Of these, eighteen concepts were related to the requirements defined in the GDPR Art.30, and the rest (29 concepts) were either supplementary to these or added by DPAs. An overview of the exercise is presented in Table 13, which shows the identified concepts and their relevance to each DPA template analysed. (Note: It has not been possible to discern a source or basis in law (EU or national) for concepts added by DPAs).

The analysis of regulator templates requires that ERoPA must be capable of expressing additional concepts beyond the concepts found in the GDPR Article 30 legal text and must be

capable of expressing all concepts found in DPA templates, as displayed in Table 13. This analysis of Regulator templates is published in the Journal of Information [18]. This work was reviewed by legal experts at conferences [13], [33] and by the expert group of the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG)[23]. This capability of expressing all concepts found in DPA templates addresses the second requirement for ERoPA (see Section 4.7 Requirement 1.2).

4.5.3 Regulator Guidance on RoPA

In this section, an analysis of regulator RoPA guidance and inspection documents is conducted to gather any requirements that need to be included within ERoPA. This qualitative analysis will involve inductive reasoning based upon knowledge gathered from the background and state-of-the-art chapters, as well as deductive reasoning to gather the key themes around RoPA best practices identified within the documents. The authoritative sources used here are guidance documents from the following regulators: the Irish Data Protection Commission (DPC) [11], The UK Information Commissioners Office (ICO) [49] and the French Regulator, Commission Nationale Informatique & Libertés (CNIL)[24]. This desk-based research also uses a guidance document that the Irish DPC issued after reviewing organisational RoPA [11]. The four regulator publications papers were selected first as they are the primary source of best practices for RoPA Maintenance and extend beyond the legal requirements and RoPA templates. These four documents are analysed to identify the key themes present. The purpose of this analysis is to identify best practices for RoPA Maintenance. These best practices are reviewed to determine whether they are suitable requirements for ERoPA. It is safe to infer that the information gathered from each template can be viewed as consistent and applicable across the EU, as the EU consistency mechanism (Art. 63) ensures that all regulatory authorities apply the GDPR consistently.

The analysis identified the following best practice themes:

- The RoPA should be divided into the different business functions within the organisation (e.g., finance, HR, marketing). The DPC recommends creating separate tables or spreadsheets for each business unit combined within an overall RoPA document[11]This leads to a requirement that ERoPA has a specification for the Interoperability of GDPR accountability information with all relevant RoPA stakeholders (see Section 4.7 Requirement 2.1).
- The UK regulator ICO suggests carrying out a data mapping exercise to determine what data the business holds and where it is used [176].This requires ERoPA to have the

[23] https://www.w3.org/community/dpvcg/
[24] https://www.cnil.fr/en/gdpr-toolkit/record-processing-activities

capacity to integrate data from heterogeneous data sources, linking data from existing GDPR accountability information sources (see Section 4.7 Requirement 1.5).

- The RoPA should Include relevant extra information where appropriate. The ICO and the French DPA CNIL[25] suggest that the RoPA should include helpful extra details not explicitly required under Article 30 of GDPR. For example, the Article 6 legal basis for processing, Article 9 basis for processing 'Special Category Data', whether a breach has occurred, and risk levels allocated to processing activities. The DPC emphasises that organisations should highlight which information is mandatory under Article 30 and which is included as a 'helpful extra' [11]. This leads to a requirement for ERoPA to be capable of the flexible inclusion of extra information (see Section 4.7 Requirement 3.3)
- The ICO recommends that RoPA be a 'living dynamic' document updated regularly [176]. The RoPA should be reviewed and updated regularly, and obsolete processing activities should be marked as such or removed from the RoPA (and archived for accountability). Hence, there is a requirement that ERoPA entries have lifecycle tracking capabilities so that the entries are traceable, and the provenance is known. This enables the organisation to know what processes are active at any time (see Section 4.7 Requirement 3.4)
- The DPC review of Organisational RoPA states that organisations must have their RoPA readily available to the DPC on request and present within ten days[11]. The requirement for ERoPA is to be capable of generating RoPA in the format of a regulator template. (see Section 4.7 Requirement 3.5)
- The DPC recommends that RoPA contain adequate detail and granularity [11] . For example, when listing the categories of personal data being processed or the technical and organisational security measures. Hence, there is a requirement for ERoPA to contain a sufficient granularity level to reflect processing activities accurately (see Section 4.7 Requirement 3.2)

The review of the guidance documents has identified many requirements for ERoPA. Regulatory requirements for RoPA extend far beyond the requirements in the legal text. Each requirement has been recorded in the ERoPA requirements listing (see Section 4.7).

[25] https://www.cnil.fr/en/record-processing-activities

4.6 Analysis of Commercial Practice to Identify Requirements for ERoPA

This section examines commercial practice to elicit the requirements for an ERoPA. This analysis consists of two data sources: commercial tools and initiatives and a survey of Data Protection Professionals.

4.6.1 Current Commercial Tools and Initiatives

Section 4.2.2 discusses the Data Privacy Industry organisations' approach to tool development for privacy automation. The research identified 364 privacy tech vendors providing software solutions covering an extensive range of Privacy services [62]. The section reviews the approach commercial tools are taking towards RoPA generation and maintainability and identifies key features present in commercial tools for RoPA. These features will be analysed to determine the requirements for ERoPA.

Automated GDPR accountability information discovery. Many Privacy software vendors offer data mapping tools to support their RoPA creation process [62]. This approach aligns with the ICO-recommended approach for a data discovery process to be conducted when creating a RoPA [49]. An analysis of privacy vendors shows that 198 vendors offer data mapping tools, thus emphasising the importance of gathering data for RoPA inclusion. The two prominent vendors [62], Big ID and One Trust propose data mapping as a first step to 'help ensure nothing is missed.' One Trust describes automated data mapping as discovering data across an organisation's entire information technology infrastructure with deep scans of data. This is achieved by utilising five hundred integrations and removing manual data discovery. Big ID proposes using its RoPA mapping App to centralise its RoPA Management Process. Commercial tool vendors suggest that organisations can build an accurate, efficient, and scalable data inventory with automated data discovery and classification that offers complete visibility into your personal and sensitive data across all types of data and all data sources. The ability to use Commercial tools to gather and classify data GDPR accountability offers significant benefits to the organisation; however, it does mean that the data is classified using proprietary ontologies rather than open-sourced vocabularies. Thus, ERoPA must manage data from heterogeneous data sources and be capable of importing information from existing GDPR accountability sources (see Section 4.7 requirement 1.5).

Interoperability to enable automated creation and monitoring of a record of processing activities. Many commercial organisations selling privacy software provide tools for importing data into the RoPA from GDPR accountability sources. These may come as simple upload Excel documents, such as the Privacy Engine spreadsheet upload tool or more complex APIs or pre-built connectors for specific enterprise applications. These import tools offer labour savings and

efficiency for onboarding RoPA data and enable the centralisation of privacy solutions without time-consuming compatibility efforts. The interoperability of RoPA is vital to providing a holistic view of the organisation's risks. Privacy companies have recognised that the interoperability of RoPA[26] This enables organisations to use their limited resources to track, monitor, and update processing records centrally. However, to ensure that GDPR accountability information can be exchanged between stakeholders, interoperability with relevant RoPA stakeholders, such as organisational units, is required (see Section 4.7 requirements 2.1, 2.2 and 3.1).

Continuous Compliance: Many Privacy vendors suggest that privacy tools reduce time spent on the manual, labour-intensive process with automated data flows. BigID[27] proposes a central RoPA that allows real-time insights whenever accountability information related to each business process is entered, updated, or identified. This automated approach enables continuous updating of RoPA before compliance becomes an issue. Hence, ERoPA must be capable of being updated at intervals when Business processes are entered, updated, or identified, thus enabling an up-to-date view of what processes are active at any time (see Section 4.7 requirement 3.4).

Using RoPA to generate other GDPR compliance outputs such as Privacy notices: As RoPA is a central repository of all the organisation's processing activities, it is a valuable resource for generating other GDPR compliance documents. Signatu It is a software-as-a-service (SaaS) platform that manages privacy on sites and apps. It utilises a processing specification to create outputs such as GDPR-compliant privacy policies [36]. Hence, the ERoPA should be capable of expressing all GDPR concepts found in all GDPR Regulator RoPA to enable the ERoPA to have the necessary GDPR concepts to generate other GDPR Accountability documents (see Section 4.7 requirement 1.2).

4.6.2 Survey of Data Protection Professionals

In this final section of the ERoPA requirements gathering, the research looks to the primary user of the RoPA, the DPO, to ascertain their requirements for ERoPA. Gathering these opinions uses an anonymous survey to collect the following information from DPOs.

- What are the most significant issues organisations are facing with RoPA?
- Who are the actors that would interoperate with RoPA?
- What are the highest priority areas that the automated RoPA could be used to address?
- What essential features should this automated RoPA system should contain to meet the needs of data protection practitioners?

The survey was broken into eight main sections. Sections 1-3 concerned the overview of the study, the plain language statement of the study, and the Informed consent, respectively.

[26]https://www.onetrust.com/resources/scaling-records-of-processing-with-data-mapping-automation-webinar/
[27] https://bigid.com/

Section 4 was used to establish the respondents' experience level to ensure they have the necessary data protection experience to be included as a valid response. This section also sought to gather the opinions of DPOs who are solely legally qualified and DPOs with technical knowledge to get a blended view. Section 5 evaluates the extent of the challenges faced by data protection professionals when maintaining RoPA. Section 6 sought to assess the role of some of the critical enablers of RegTech that may need to be in place for automated interoperable GDPR tools to become established. The respondents were provided with several statements and asked how much they agreed with them. Section 7 clarifies where and how a respondent might use ERoPA and seeks to establish the importance of some key features of the ERoPA.

The survey was piloted and validated by three practising Data Protection peers of the researcher. The survey uses guidelines for research in generating the questions set for the study to ensure that the best practices of content, construct, and reliability are met. The survey received ethics approval from the Dublin City University Ethics Committee (see Appendix A). The survey was conducted between 26th Oct 2022 and 22nd Nov 2022. Participants were gathered using DPO peer networks and LinkedIn and collected from industry, consultancy, and academia. The respondents were provided with an overview of ERoPA and the study's objectives in section 1. All participants were provided with an explanation of ERoPA and then asked questions to complete the survey using Google Forms. A total of forty-four responses were received. Of these, only 43 consented to participate; therefore, the sample size was n=43.

The responses were gathered and validated using the questions in section 4 of the survey to ensure the respondents met the DPO or legal qualification criteria. Of n=43 participants, 34 of the 43 had more than three years of experience in the Data Protection / Privacy domain. Eight participants had 1-3 years of data protection/ privacy domain experience, and only one respondent had less than one year of experience, but that person had gained a data protection qualification. This indicated all respondents had relevant experience in data protection.

The review of the respondents in section 4 also showed that the respondents hold a high level of qualification, as per Table 14 below. This demonstrates that the sample has a sufficient level of expertise with a blend of legal and technological experience.

Table 14 Qualifications of Respondents to Survey.

Qualifications	**Number of respondents n=43 (one respondent did not consent to use of data)**
Data protection certificate, diploma, degree or similar, e.g. CIPP[28]	20
Information Technology / Computer Science Degree or greater, Data protection certificate, diploma, degree or similar, e.g. CIPP	13
Qualified Lawyer or Solicitor	4
Qualified Lawyer or Solicitor, Data protection certificate, diploma, degree or similar, e.g. CIPP	3
Information Technology / Computer Science Degree or greater	2

Of the forty-three respondents, twenty-four hold or previously held the role of Data Protection Officer for an organisation, whilst the other nineteen did not. This indicates that whilst the sample contains a high level of qualification, not all respondents are DPOs, so it is best to refer to the respondents en masse as data protection professionals.

Section 5 of the survey[29] seeks to evaluate organisations' existing challenges in maintaining RoPA (see Appendix A). The survey reinforced previous research verifying that DPOs face significant challenges with RoPA maintenance [4,5]. The scale of these challenges is evident from the survey, where only 10% agree or strongly agree that their RoPA is accurate. Concerning organisational buy-in to RoPA, only 20% of organisations agree or strongly agree that their organisation is bought into RoPA. Similarly, only a small number of respondents (20%) agree or strongly agree that their RoPA is up to date, and only 38% of organisations agree or strongly agree that their RoPA is comprehensive.

The survey identifies that many respondents report a lack of commitment to RoPA in their organisation (60 %) (see Figure 9).

[28] https://iapp.org/certify/cippe/

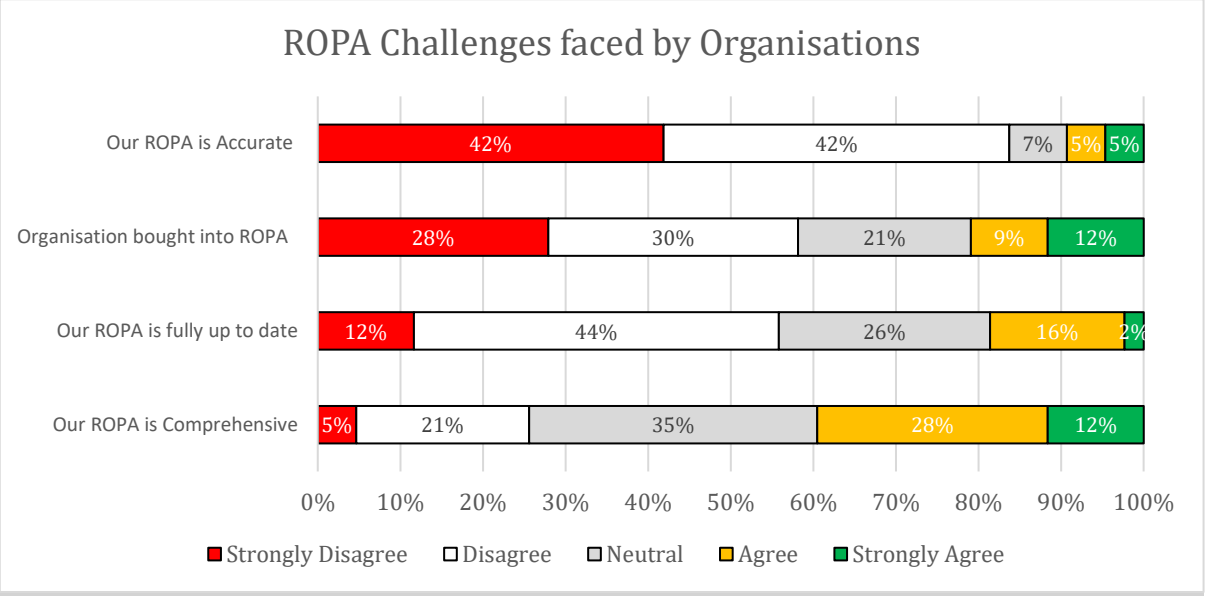

Figure 9 Survey Results - How challenged are organisations with Maintaining RoPA?

Section 6 of the survey seeks to understand how the critical enablers of GDPR RegTech are to be in place for automated interoperable GDPR tools to become established. The DP professionals were asked to what extent respondents agreed with these statements.

- *The adoption of developments in technologies would help with the automation of GDPR compliance.*
- *Automated interoperable GDPR tools will only happen if the Data Protection Supervisory Authorities drive adoption.*
- *For automated GDPR compliance tools to be developed, there will need to be agreement on common standards/ agreed semantics (definitions of terms) for personal data processing.*
- *A robust data governance platform within an organisation would enable the development of automated GDPR compliance tools.*

The survey results show that respondents agree or strongly agree that adopting technological developments would help with the automation of GDPR compliance thirty-four respondents 77.3%. There was a less clearly defined result when respondents were asked if Automated interoperable GDPR tools will only happen if the Data Protection Supervisory Authorities drive adoption, in that only twenty-two respondents or 50% agree or strongly agree. Respondents felt that for automated GDPR compliance tools to be developed, there would need to be agreement on common standards/ agreed semantics (definitions of terms) for personal data processing (35 or 79.5% agree or strongly agree) and that a robust data governance platform within an organisation would enable the development of automated GDPR compliance tools 36 respondents or 81.8%.

Section 7 of the survey sought how data protection professionals would use ERoPA. The E ERoPA system would allow data processing information to be automatically exchanged/shared and inspected between data processing actors. These actors could include internal organisational units/ departments, data processors, data controllers, auditors, regulators, or the data protection officer. These parties all require access to, sharing or input to RoPA records. The survey sought to establish which of these relationships would benefit the most from ERoPA. This aimed to develop an understanding of the priority of these relationships, to understand if sharing was required with all actors, and if some actors' exchanges were highly represented, indicating that these exchanges between parties were of a higher priority.

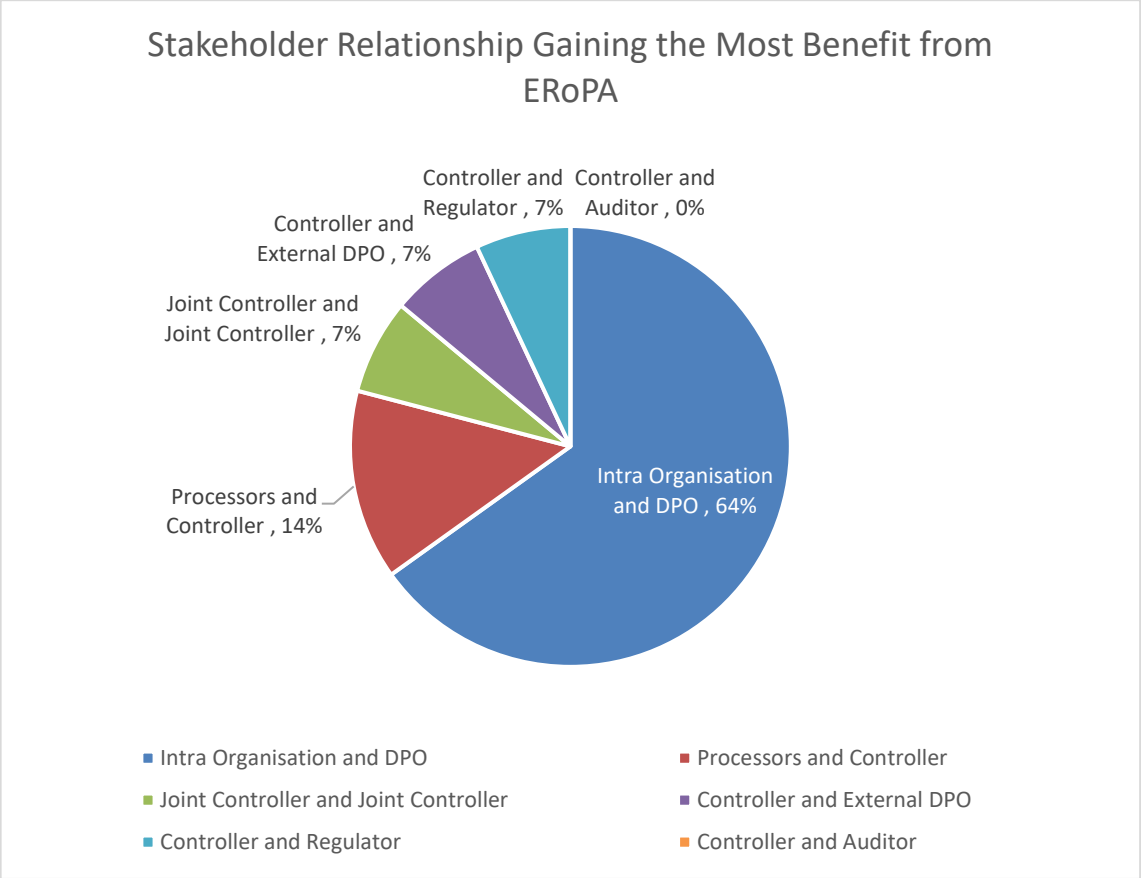

Figure 10 Stakeholder Relationship Gaining Most Benefit from ERoPA.

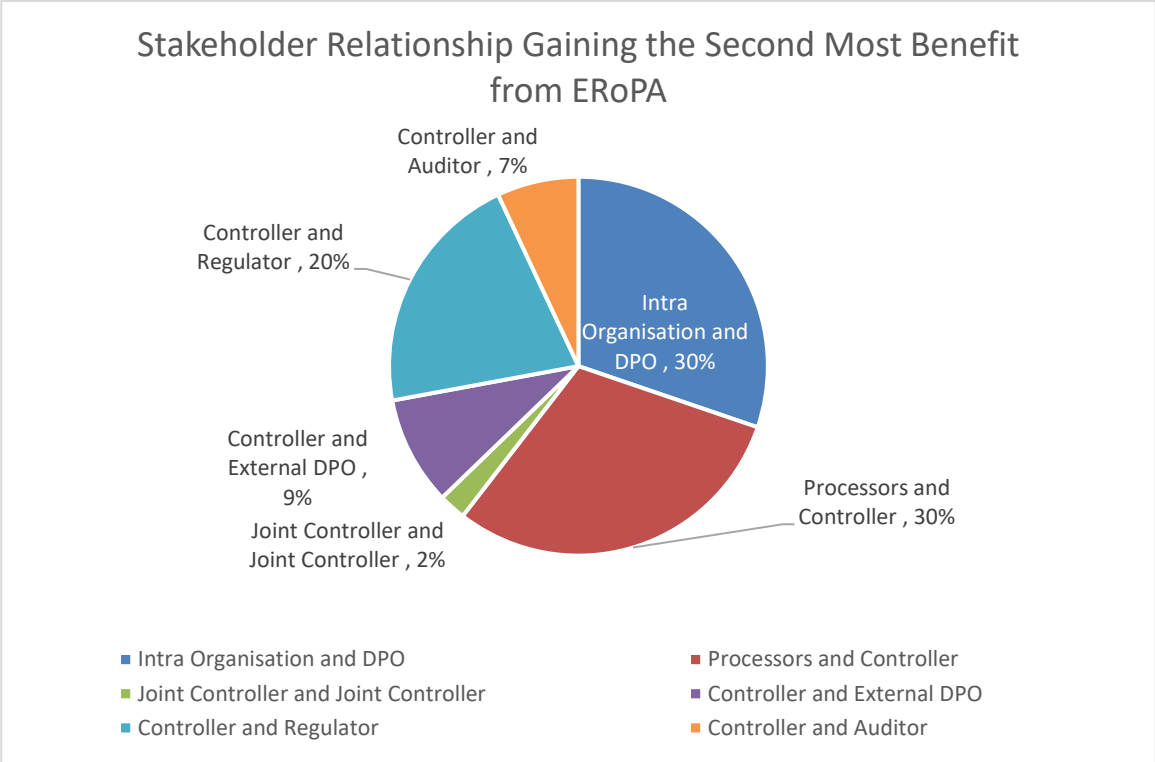

Figure 11 Stakeholder Relationship Gaining the Second Most Benefit from ERoPA.

The survey results indicate which actor relationship will benefit the most from ERoPA. Several information flows interact with RoPA [176]. These information flows, such as organisational and controller/ DPO, can be internal. Data may flow between external processors and the Controller. There are also several extra organisation data flows between the Controller and another controller, an external DPO, certification body or regulator. The respondents were offered these six choices of GDPR interoperability Actor relationships (see Figure 10). The respondents were asked to rank this relationship in terms of which of the six would benefit most from ERoPA. This question aimed to identify which critical actor interoperability areas the research should be focused on. The respondents indicated that their primary usage would be between the Intra-Organisation units exchanging GDPR accountability information with the DPO. This data flow was selected by 29 of the n=43 respondents, ranking it as an essential data flow (see Figure 10). The second most crucial interoperability flow was between the Controller and Processor. 20 of 44 DPOs identified this flow within their top two choices (see Figure 11). An interesting observation is the data flow from Controller to Auditor and Controller to the regulator, which featured in the second or third choice for 18 of 44 places. This finding indicates that these data flows to certification and regulation bodies are becoming more critical. The emergence of this requirement may be prompted by the approval of the first GDPR certification bodies [177].

Section 7 also suggests ten different GDPR accountability tasks where ERoPA might be used. The DPOs were asked which accountability tasks they would use ERoPA for. The respondents could select as many responses as they wished. The respondents were prompted with a choice of ten uses for ERoPA. They were asked to choose as many uses as they considered ERoPA could be used for. In this question, the research aimed to understand which uses of ERoPA were the most prevalent and which were less prevalent, thus allowing the researcher to focus on the priority areas. In general, the respondents wish to use ERoPA across a broad spectrum of uses, with the most prevalent use case being to identify gaps, conflicts, or non-compliance with RoPA. This finding indicates the potential of ERoPA to assist the DPO across a broad spectrum of accountability tasks. These tasks are displayed in Figure 12.

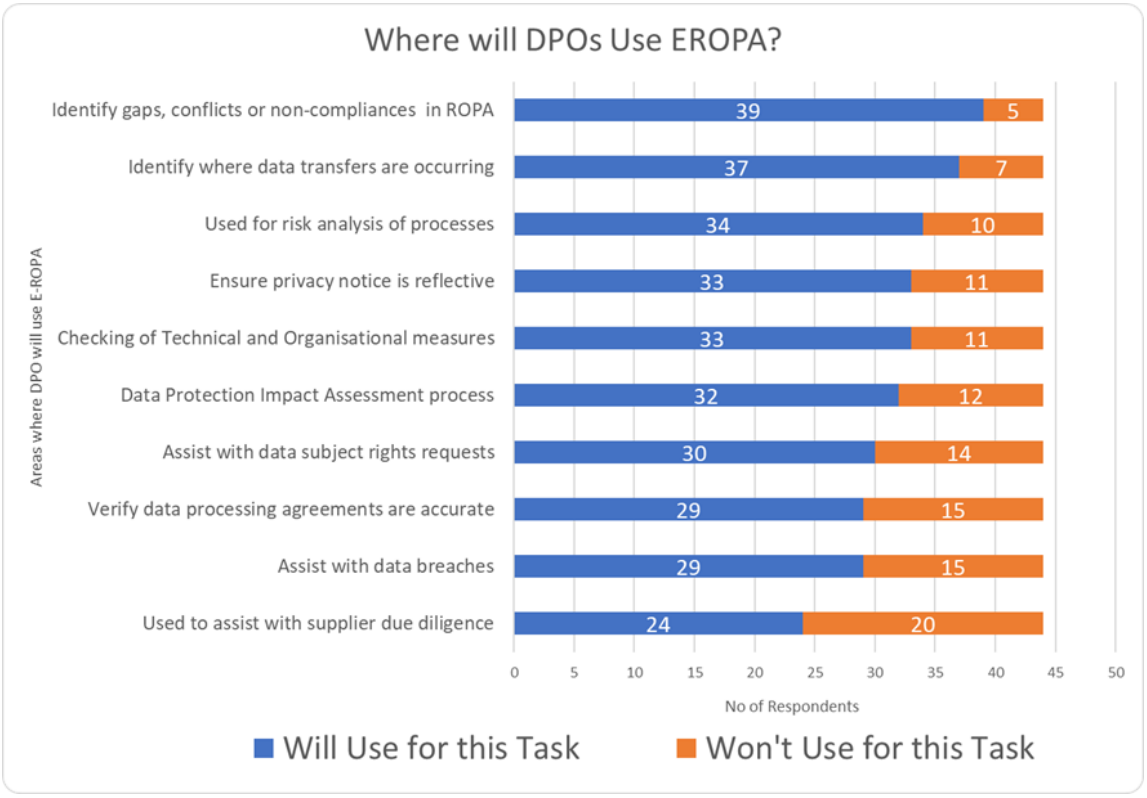

Figure 12 Where will the DPO Use ERoPA?

Discussion on Survey: The survey results provided perspective on the challenges facing DPOs, leading to several ERoPA requirements. It also confirmed the motivation for this research in that the respondents confirmed that RoPA needed to be more accurate and up to date in their organisations. The respondent confirmed that a robust data governance platform within an organisation would enable the development of automated GDPR compliance tools, and they also confirmed that for automated GDPR compliance tools to be developed, there would need to be agreement on common standards/ agreed semantics (definitions of terms) for personal data processing. The respondents indicated that there were many actors with whom they would

exchange GDPR accountability and many GDPR accountability tasks with which they would use RoPA. This has led to the following requirements for ERoPA combined into a common specification in the next section.

1. The survey indicated ERoPA would be used for many GDPR accountability tasks beyond RoPA. Hence, the RoPA must use an open, extensible vocabulary with all information concepts found in the GDPR domain to enable the automation of regulatory compliance processes (see Figure 12 for examples of the regulatory tasks) (see Section 4.7 Requirement 1.3).
2. The Survey identified that the stakeholders require ERoPA to be capable of exchanging GDPR accountability information with other key stakeholders. Best practice from GDPR Regtech would indicate that this is best achieved with a standardised semantic approach[32]. Hence, the ERoPA should provide a specification for enabling Semantic Interoperability of GDPR accountability information with all relevant RoPA stakeholders [18](see Section 4.7 Requirement **2.1.)**
3. The ERoPA should use an agreed-upon interoperability specification for the representation, collection, and transfer of RoPA. This interoperability specification should contain all information concepts in the RoPA domain. The Interoperability specification will be best served with a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues [32] , as this would encourage the adoption of the specification (see Section 4.7 Requirement 2.2)
4. The survey indicated that a critical requirement of ERoPA was the identification of compliance gaps and conflicts in the RoPA. The ERoPA should support testing for compliance and automated risk identification to assist the DPO in identifying non-compliances. (see Section 4.7 Requirement 3.6) and the Ontology format must support queries to enable the DPO to analyse the model (see Section 4.7 Requirement 1.6)

In the next section, the Requirements gathered from section 4.5 are combined into a specification of the requirements for ERoPA.

4.7 Specification of the Requirements for ERoPA Approach

In Section 4.3, an analysis of the requirements for a machine-readable RoPA is conducted. In this section, these requirements are grouped into three main categories:

- A semantic ontology for the **representation** of GDPR Accountability concepts necessary to create and maintain RoPA (see Section 2.3.2).

- An enabling semantic interoperability specification for the **collection** and **transfer** of RoPA information from stakeholders such as organisational units or data processors and relevant data protection stakeholders (see Section 2.3.3).
- Application requirements to enable the validity and conformance of RoPA and provide DPOs with tools for the **review** and **inspection of** RoPA in both human-readable and machine-readable formats (see Sections 2.3.3 and 2.3.4).

The requirements for each category of ERoPA to enable GDPR compliance automation are gathered and presented below. Each requirement is provided with a unique number to track each requirement through the evaluation process of this research in this thesis. A traceability matrix for each of these requirements is presented in Table 15 in Section 4.8.

The first category of requirements concerns a semantic ontology for the **representation** of GDPR Accountability concepts necessary to express RoPA:

- **Requirement 1.1** The machine-readable RoPA must be capable of expressing all information concepts found in GDPR Article 30 [30]
- **Requirement 1.2** The machine-readable RoPA should be capable of expressing all GDPR concepts found in all GDPR Regulator RoPA templates [18].
- **Requirement 1.3** The RoPA must utilise an open, extensible vocabulary with all concepts in the GDPR domain [17], [32] to support systems to exchange information seamlessly and to enable consistency and standardisation of terms to support regulatory compliance.
- **Requirement 1.4** The RoPA must utilise a standardised vocabulary where GDPR concepts are agreed upon by technologists and legal experts [20], [41], [115].
- **Requirement 1.5** The ERoPA must be able to integrate data from heterogeneous data sources and link data from existing GDPR accountability information sources [18].
- **Requirement 1.6** The machine-readable representation of information must support queries when used to model an organisation's RoPA to enable the DPO to analyse the model of the organisation's RoPA.

The second category of requirements concerns a specification for enabling Semantic Interoperability, **collection** and **transfer** of GDPR accountability information between different data protection stakeholders:

- **Requirement 2.1** The machine-readable RoPA must have a specification for the Interoperability of GDPR accountability data with all the following relevant RoPA stakeholders [18]: (e.g. controllers, processors, organisational units, regulators and auditors/certification bodies).

[30] GDPR Art 30

- **Requirement 2.2** The machine-readable RoPA should utilise an agreed-upon interoperability standard specification for the representation, collection and transfer of RoPA information in a machine-readable and interoperable manner to describe datasets so that publishers can use a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues [17], [18], [32].

The third category of requirements concerns Application requirements to enable the validity and conformance of RoPA and provide DPOs with tools for **review** and **inspection** of RoPA in both human-readable and machine-readable formats (see Sections 3.3.3 and 3.3.4).

:

- **Requirement 3.1** The organisation requires ERoPA to enable communication of GDPR accountability information into RoPA, where it can be viewed and analysed by the organisation's DPO [49].
- **Requirement 3.2** The ERoPA must contain a sufficient granularity level to reflect processing activities accurately [11].
- **Requirement 3.3** The ERoPA must be capable of flexible inclusion of extra information.
- **Requirement 3.4** There is a requirement that ERoPA entries must have lifecycle tracking capabilities so that the entries are traceable, and the provenance is known. This enables the organisation to know what processes are active at any time [11].
- **Requirement 3.5** The RoPA must be available upon request for submission to a GDPR regulator in the format of a regulator template. This will require the ERoPA to be able to generate a RoPA in the required template (see GDPR Recital 82).
- **Requirement 3.6** The ERoPA should support non-compliance testing and automated risk identification to assist the DPO in identifying non-compliances.

4.8 Traceability Matrix of ERoPA Requirements

This chapter uses a requirement engineering approach to identify a set of ERoPA requirements gathered from six sources. The combination of established sources such as primary research, academic publications and industry practice ensures that a comprehensive, rigorous approach has been taken to gather the requirements. The researcher's experience as a practising Data Protection officer enables a perspective from day-to-day industry practice to validate the requirements. Table 15 provides a summary of the data sources and requirements.

Table 15 Requirements Traceability Matrix.

Requirement No.	Requirement Description	Source of Requirement				
		Art 30 / GDPR text	Guidance Documents	SoA review	Industry Practices	Survey of DPOs
1.1	Expressing all information concepts in Article 30	✓				
1.2	Expressing all GDPR concepts in Regulator RoPA templates		✓			
1.3	Use an open, extensible vocabulary.			✓		
1.4	GDPR concepts are agreed upon by technologists and legal experts			✓		
1.5	Integrate data from heterogeneous data sources.			✓		
1.6	Must support queries to enable the DPO to analyse the RoPA.					✓
2.1	A specification for the Interoperability of GDPR accountability data with all relevant RoPA stakeholders					✓
2.2	Use an agreed-upon interoperability standard specification.			✓		
3.1	Enable communication of GDPR accountability information into RoPA					✓
3.2	Contain sufficient granularity level to reflect processing activities accurately.		✓			
3.3	The ERoPA must be capable of flexible inclusion of extra information.		✓			
3.4	Entries are traceable, and the provenance is known.		✓			
3.5	RoPA is available on request for submission to a GDPR regulator.	✓				
3.6	Support non-compliance testing and automated risk identification.					✓

4.9 Summary of ERoPA Requirements Analysis

This chapter conducts a systematic analysis to identify ERoPA requirements. The research gathers the requirements from six primary sources: legal texts, academic sources, and practice. These

sources are analysed using desk research to identify requirements from the legal texts and regulator templates. An analysis of the scholarly research in GDPR RoPA compliance is conducted, a study of commercial tools for RoPA is undertaken, and a survey of Data protection Professionals is completed. The results of this analysis identify fourteen requirements for ERoPA. The requirements are categorised and presented in section 4.7. In the next chapter, the design of the ERoPA is presented.

5 The Common Semantic Model of RoPA (CSM-RoPA)

5.1 Chapter Overview

This chapter details creating and evaluating the Common Semantic Model of RoPA (CSM-RoPA) ontology. Using the ADR methodology, this forms the first build, implement, and evaluate (BIE) stage of ERoPA, as illustrated in Figure 13. CSM-RoPA addresses the research sub question RSQ2.b, which is to develop an ontology to implement the CSM-RoPA using Semantic Web standards and best practices.

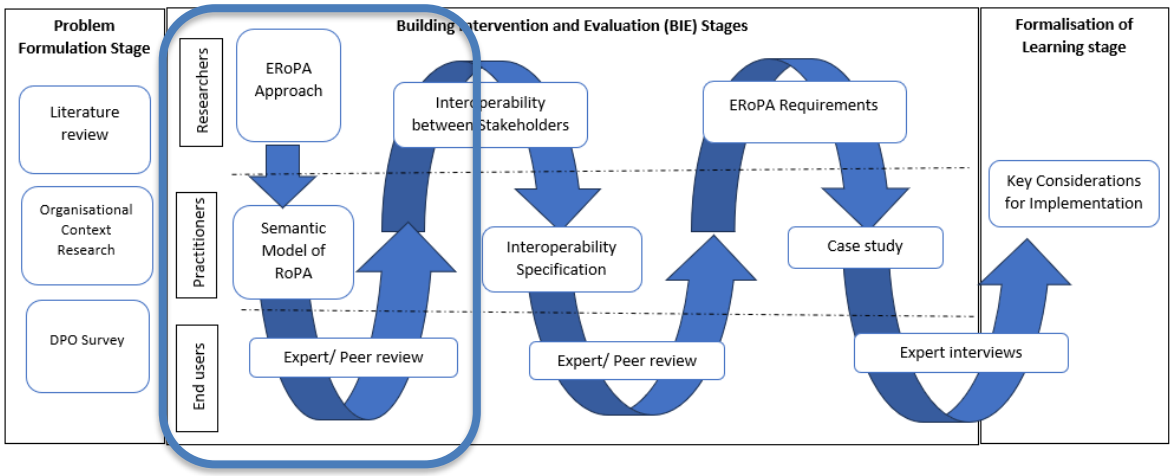

Figure 13 First Build, Implement and Evaluate Stage of ERoPA Development.

The chapter introduces the development and evaluation of the CSM-RoPA ontology, which is essential for implementing the machine-readable RoPA approach. This involves creating a semantic model of RoPA based on requirements collected from GDPR and the Data Protection Authorities. Section 5.2 describes the design and specification for CSM-RoPA, and the ontology engineering methodology used; Section 5.3 presents the implementation of the ontology, and Section 5.4 evaluates the ontology. In Section 5.5, the learnings from the first BIE stage are summarised and how they will impact future stages/ implementation.

5.1.1 ADR Roles

The development of CSM-RoPA utilises designated ADR Roles to support its development as a component of the ERoPA Approach (see Table 16). The first of these roles, the researcher who is a practising DPO with experience of conducting the planning and implementation of the ERoPA, conceptualises the design for CSM-RoPA, while the second role, the Practitioner role, is met by the

Data Privacy Vocabularies and Controls Community Group (DPVCG)[31], who contribute their legal and technical expertise and context to the development and implementation of the artefact. The DPVCG comprises technologists, researchers and legal experts who produce the Data Privacy Vocabulary (DPV) as a deliverable. The third ADR role, " users" for the CSM-RoPA ontology are formed of (i) peer-reviewed publications where experts provide their feedback (ii) industry usage by organisations, and reference to the (iii) original DPO survey requirements.

Table 16 ADR Role Assignment for BIE1

Role	Assignment
Researcher	The Thesis Researcher (a practising DPO)
Practitioner	Data Privacy Vocabularies and Controls Community Group (DPVCG)
User	Peer-reviewed publications, Industry usage, Original DPO survey requirements

5.2 CSM-RoPA Design

Based on the State of the Art (Chapter 3) and the requirements for ERoPA (Chapter 4), an ontology called the common semantic model of RoPA (CSM-RoPA) has been designed. CSM-RoPA forms a vital component of ERoPA as it represents the GDPR concepts and information essential for RoPA in a machine-readable form. CSM-RoPA was created based on a comprehensive analysis of the GDPR Article 30 legal text (see Section 4.5.1) and a set of seventeen regulator-supplied RoPA templates (see Section 4.5.2). CSM-RoPA is thus integral to supporting the functioning of ERoPA toolchains across organisations and jurisdictions.

5.2.1 Ontology Engineering Methodology

For developing the Common Semantic Model of RoPA (CSM-RoPA) in the first BIE stage, the thesis utilises the NeOn (Networked Ontologies) methodology [38]. This same approach is also employed by the Data Privacy Vocabulary, which will be used, extended, and contributed to as part of this thesis (see Section 1.7).

The Neon methodology is chosen as it is a comprehensive framework for the collaborative development of networked ontologies. It provides guidelines and practices for ontology development, emphasising the reuse and integration of existing ontologies and data sources. There

[31] https://www.w3.org/groups/cg/dpvcg/

are several key benefits of the NeOn Methodology. NeOn offers efficiency by reducing the time and effort required to develop ontologies and promoting reuse and collaboration. Neon provides a flexible approach that supports various ontology development scenarios and contexts. NeOn is also suitable for creating large-scale, interconnected ontology networks, offering the opportunity to extend CSM-RoPA beyond the domain of GDPR accountability data solely for RoPA. NeOn also enhances the quality and consistency of ontologies through systematic validation and evaluation. By following the NeOn Methodology, organisations can create robust and interoperable ontologies that effectively support their knowledge management and Semantic Web applications. The NeOn Methodology guides different aspects of the development of an ontology. Each scenario responds to specific activities such as requirements gathering, ontology reuse, ontology engineering, design enrichment and validation. Table 17 presents the steps followed to create the CSM-RoPA Ontology.

Table 17 CSM-RoPA Ontology Engineering Workflow Steps.

Steps	**NeOn Methodology Stage**	**Task**
1	Specify the requirements for the model - define the goals, scope, and requirements for the ontology.	Identify requirements from the analysis of State of the Art (see Section 4.7)
2	Knowledge acquisition - gathering of domain knowledge	Model concepts from GDPR Article 30 (see Section 4.5.1) and regulator-supplied RoPA templates (see Section 4.5.2)
3	Ontology reuse and reengineering - reuse existing ontologies or ontology modules, modifying them to fit the current needs	Identify reuse of existing terms from the Data Privacy Vocabulary.
4	Ontology design - develop the ontology, specifying classes, properties, instances, and axioms.	Create additional terms to express missing RoPA concepts.
5	Implementation - translate the ontology design into an actual implementation in an ontology language	Implement ontology using RDFS, SKOS and OWL for extending the DPV. Generate documentation using WIDOCO.
6	Evaluation - evaluate the ontology to ensure it meets the requirements and is of high quality	Evaluate the extent to which the ontology satisfies competency questions, follows best practices, and assesses its extent to express a RoPA. Validate it through peer review.

The structured approach of the NeOn methodology provides a systematic and disciplined approach to ontology engineering, ensuring the development of high-quality, reusable, and maintainable ontologies.

5.2.2 Requirement Specification for CSM-RoPA

This section presents the ontology requirements specification, which is the first stage of the NeOn Ontology engineering methodology (see Section 1.4) [38]. The ontology requirement specification design is based on the NeOn methodology template [178].

Purpose: The ontology must represent all GDPR concepts expressed as classes and relations in regulator RoPA templates. CSM-RoPA must support the DPO's obligations to monitor GDPR compliance to identify risk and non-compliance (see Section 4.7 requirements 1.1, 1.2, and 1.5).

Scope The ontology scope is all GDPR information included in GDPR Article 30 RoPA (see Section 4.7 requirement 1.1) or a RoPA template supplied by a regulator (see Section 4.7 requirement 1.2). The ontology should contain sufficient granularity to express the level of detail necessary for the organisation to meet its RoPA obligations (see Section 4.7 requirement 3.2). The ontology does not require modelling GDPR concepts not required to represent a RoPA.

Intended users: The users of the ontology will be the RoPA Stakeholders (see Section 4.7 requirements 2.1). These are as follows: (i) DPOs (ii) organisational units (iii) data controllers, (iv) data processors, (v) certification bodies and auditors and (vi) regulators.

Intended use: The ontology must support the DPO in completing their GDPR compliance tasks as per GDPR article 39, summarised below:

1. To inform and advise the controller or the processor and the employees who conduct processing of their obligations pursuant to this Regulation and other Union or Member State data protection provisions.
2. To monitor compliance with this Regulation, with other Union or Member State data protection provisions and with the policies of the controller or processor regarding the protection of personal data, including the assignment of responsibilities, awareness-raising and training of staff involved in processing operations, and the related audits.
3. To provide advice where requested as regards the data protection impact assessment and monitor its performance under Article 35.
4. To cooperate with the supervisory authority.
5. To act as the contact point for the supervisory authority on issues relating to processing, including the prior consultation referred to in Article 36, and to consult, where appropriate, about any other matter.

The ontology must support the controller and processor in maintaining their Article 30 obligations:

- The record must contain all details required in Article 30 (see Section 2.3 for Article 30 detailed requirements).
- The record shall be in writing, including in electronic form.

- The controller or the processor and, where applicable, the controller's or the processor's representative shall make the record available to the supervisory authority on request (see Section 4.7 requirements 3.1, 3.2, 3.4, 3.5).

Ontology requirements: The ontology should support the following non-functional requirements. The ontology should be documented in English and aligned with a standardised, interoperable standardised vocabulary. The ontology must ensure that terms utilised for inclusion are agreed upon between technologists and legal [41]. The ontology should support answering the following competency questions (see Table 18), which are based upon the GDPR Article 30 requirements concerning the Record of Processing Activities.

Table 18 Competence Questions that CSM-RoPA must meet based on GDPR Art.30.

Question No.	Competence Question (based on text taken from GDPR Article 30)
CQ01	What is the purpose of the data processing activity?
CQ02	What categories of personal data are being processed?
CQ03	Who is the data controller responsible for the processing activity?
CQ04	Who are the data processors involved in the processing activity?
CQ05	What is the legal basis for the processing (e.g., consent, contract, legal obligation)?
CQ06	What is the duration or retention period for the data being processed?
CQ07	Who are the data subjects whose personal data is being processed?
CQ08	What are the data subjects' rights in this processing activity?
CQ09	Are there any third parties or recipients with whom the data is shared?
CQ10	Is the personal data transferred to countries outside the European Economic Area (EEA)?
CQ11	What safeguards are in place for data transfers to third countries?
CQ12	What security measures are applied to protect personal data during processing?
CQ13	What are the risks associated with the data processing activities?
CQ14	Has a Data Protection Impact Assessment (DPIA) been conducted for this processing activity?
CQ15	What is the source of the data being processed?
CQ16	Are any automated decision-making processes involved in this data-processing activity?
CQ17	What are the retention periods for the data being processed?
CQ18	Who is the Data Protection Officer (DPO) overseeing this processing activity?

Pre-glossary of terms: A review of the seventeen regulator-supplied RoPA templates (see Section 4.5.2) identified the following terms and their frequencies, presented in Table 19. The goal of this task is to extract from the list of CQs a pre-glossary to be used in the conceptualisation activity. Each term is mapped against the CSM-RoPA competence question that must be met, e.g. the term 'Recipient categories' is met by CQ09, which is 'Are there any third parties or recipients with whom the data is shared?'. When terms that do not have a competence question are identified, they are marked as follows (-), and an additional competence question is formulated below.

Table 19 Terms Identified in Regulator Templates and Frequency of Occurrence.

Term	Frequency	Term	Frequency
Third Countries in Data Transfer (CQ10) (CQ11)	17	Processing Status (-)	4
Tech/Org measures (-)	17	Data Subject Rights (CQ08)	4
Representative Contact (-)	17	Data Sources (CQ15)	4
Representative (-)	17	Third-Party Data Transfer (CQ09)	3
Recipient categories (CQ09)	17	Vulnerable Data Subject Category (-)	2
Purposes of processing (CQ01)	17	Type of Processing (-)	2
Personal Data Categories (CQ02)	17	Relevant DPIA (CQ14)	2
Nature of Transfer (-)	17	Record of consent (-)	2
Legal basis (CQ05)	17	Legitimate interests (CQ05)	2
Joint Controller Name (-)	17	Joint Controller agreement (-)	2
Joint Controller contact (-)	17	Impact Assessment, Prior Consultation (-)	2
Data Subject Categories (CQ07)	17	Data Breach (-)	2
Data Protection Officer (CQ18)	17	Automated decision-making, profiling (CQ16)	2
Data Protection Officer Contact (-)	17	Technologies used (-)	1
Data Controller (CQ03)	17	Security measures (CQ12)	1
Data Controller Contact (-)	17	Risk assessment and mitigation (CQ13)	1
Appropriate Safeguards (CQ11)	17	Personal Data Location (-)	1
Retention/Deletion Periods (CQ17)	16	Main/Auxiliary Processing (-)	1
Data Processing Contract (-)	7	Legitimate interests' assessment (-)	1
Business Process (-)	7	External DPO organisation (-)	1
Owner of Process (-)	5	Data Processors (CQ04)	1
DPIA Results (CQ14)	5	Data Combination (-)	1
Special Personal Data Category (-)	4		

Additional competency questions to support regulator RoPA templates: The review of the regulator templates identified twenty-five terms not met by the competency questions from GDPR article 30. To overcome these additional terms required by regulator templates (see Section 4.5.2), twenty-five additional competency questions are presented in Table 20.

Terms for answers: The regulator templates do not contain proposed answers, so it is not possible to define this currently.

Objects: The Ontology must support the creation of a GDPR Article 30 Record of Processing Activities and generate outputs in the format of Regulator-supplied RoPA templates. These objects are required to meet GDPR regulatory requirements.

Table 20 Supplementary Competence Questions gathered from Regulator Templates

Question No.	Competence Question
CQ19	What technical and organisational measures of security are in place?
CQ20	What are the contact details of the data controller's representative?
CQ21	Who is the representative of the data controller?
CQ22	What is the nature of the personal data transfer?
CQ23	Who are the Joint Controllers involved in the processing activity?
CQ24	What are the contact details of the joint controller?
CQ25	What are the contact details of the data protection officer?
CQ26	What are the contact details of the data controller?
CQ27	Where is the data processing contract located?
CQ28	What is the name or identifier of the business process?
CQ29	Who is the owner of the processing activity?
CQ30	What categories of special personal data are processed in the processing activity?
CQ31	What is the status of the processing activity?
CQ32	What categories of vulnerable data subjects are processed in the processing activity?
CQ33	What is the type of processing?
CQ34	Where is the record of consent located?
CQ35	Where is the joint controller agreement located?
CQ36	Has there been an impact assessment/ prior consultation for this processing activity?
CQ37	Has a personal data breach occurred related to this processing activity?
CQ38	What System or software is used (technologies used)?
CQ39	What technical and organisational measures of security are in place?
CQ40	Where is the personal data located (to support data subject rights requests)?
CQ41	Is this processing activity the organisation's main or auxiliary processing activity (distinguishing between main/core and auxiliary/secondary operations)?
CQ42	Where is the Legitimate interest assessment located?
CQ43	Who is the external entity acting as the organisation's DPO?
CQ44	Have multiple data sets been combined for the processing activity?

5.2.3 Knowledge Acquisition

This section describes the knowledge acquisition step of the NeOn methodology. This step involves collecting domain knowledge from five sources listed in Table 21.

Table 21 Data Sources for CSM-RoPA Knowledge Acquisition Stage.

Data Source	Source
GDPR RoPA legal requirements	Chapters 3.3 and 4.5
Regulator RoPA guidance documents	Chapters 3.3 and 4.5
Regulator RoPA template requirements	Chapter 4.5
Data Privacy Vocabulary	Chapter 3.5
Requirements for a machine-readable RoPA	Chapter 4.7

The gathering of concepts required for RoPA consisted of two cycles. The first of these cycles occurred in 2020 when six English-language RoPAs were gathered and analysed to identify the GDPR concepts present [13] EU Data Protection Regulators provided the six ROPA templates from the jurisdictions of Belgium, Cyprus, Denmark, Finland, Luxembourg, and the United Kingdom based on their use of the English language. The templates were analysed to identify the GDPR concepts present. A systematic review of the concepts was conducted to establish synonyms, overlapping concepts, and related concepts. The exercise yielded forty-three unique concepts representing information to be recorded in a ROPA. Of these, eighteen concepts were related to the requirements defined in the GDPR Art.30, and the rest (25 concepts) were either supplementary to these or added by DPAs. Based on the interpretation of the GDPR and the use of concepts in ROPA, direct relationships were made, such as composition or qualifications, and establishing domain and range, which were implicit in the templates These fields were consolidated into a UML model, as represented in Figure 14[32] . This UML model has since been published in a peer-reviewed publication [13] and presented at the International Conference on Legal Knowledge and Information Systems (Jurix) in 2020.

[32] The notation is standard UML and should be interpreted as follows: boxes are classes, boxed arrows represent subclasses, and general arrows represent a relationship (e.g. contains).

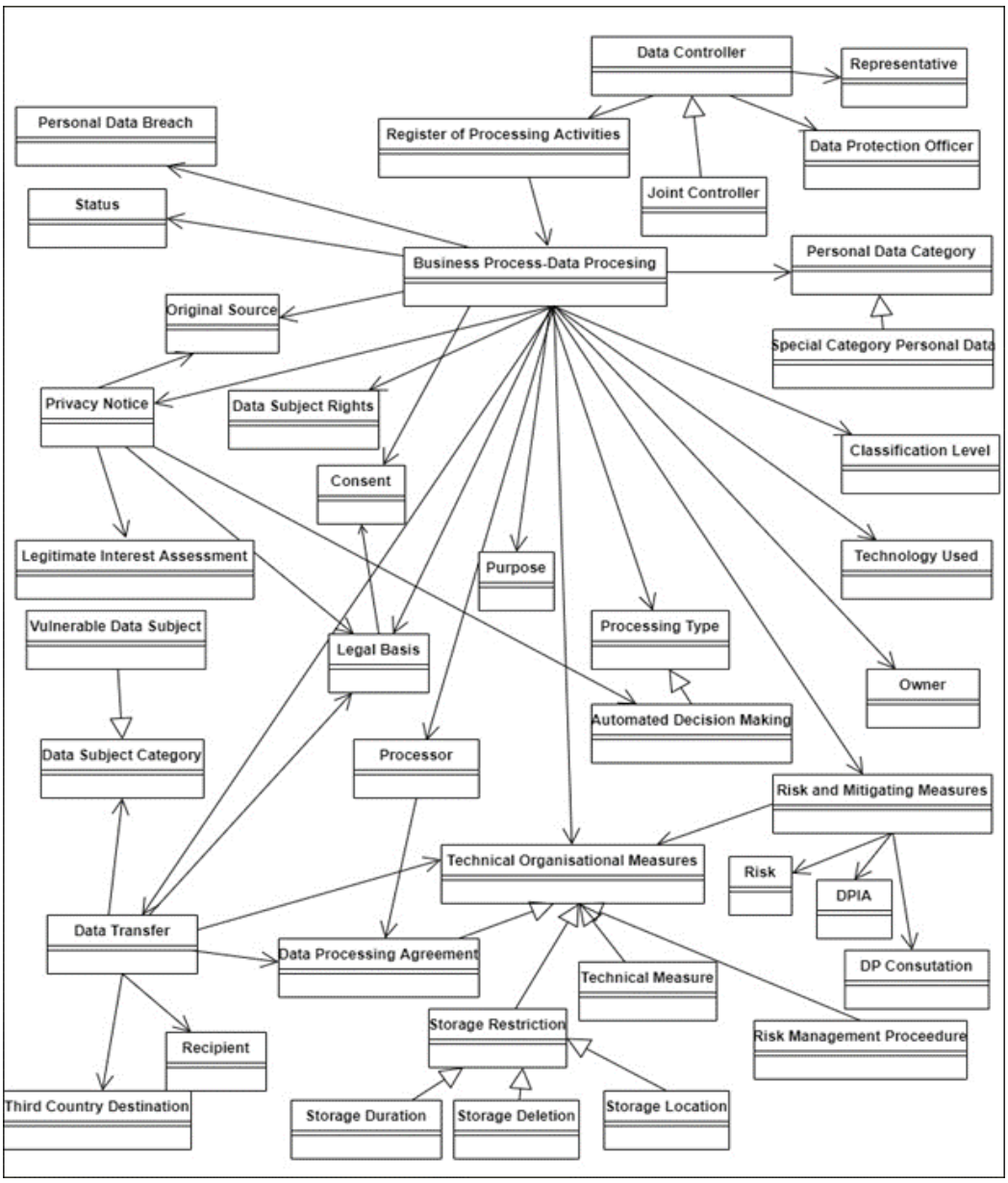

Figure 14 UML Representation of the ROPA Model based on Regulator Templates.

A second more comprehensive and extensive RoPA gathering exercise occurred in January 2022 [18], which expanded the analysis to 31 DPAs' and found that 17 DPAs provided ROPA templates in a language other than English. To analyse these, the templates were translated to English using Google Translate service and manually verified by using the author's domain expertise and experience[33]. Like the first cycle, a systematic review of the concepts was conducted to establish synonyms, overlapping concepts, and related concepts. The exercise yielded four additional concepts, bringing the total of GDPR concepts to forty-seven unique concepts for representing information to be recorded in a RoPA. Of these forty-seven concepts, eighteen were related to the requirements defined in the GDPR Art.30, and the rest (29) were either supplementary to these or added by DPAs.

[33] The researcher is a certified Data Protection Officer

5.2.4 Ontology Reuse and Reengineering

The previous section describes gathering GDPR concepts from the regulator RoPA templates. The first of these cycles identified forty-three concepts, with the second cycle identifying forty-seven concepts. This section will detail the process followed to represent these GDPR concepts semantically.

The CSM-RoPA ontology incorporates concepts from the Data Privacy Vocabulary (DPV) to enhance semantic modelling. The DPV is an ideal vocabulary for reuse, as it is widely recognised and designed to standardise and simplify the description of data privacy concepts, regulations, and policies in a machine-readable format (see Section 2.5.4). It offers a framework for representing and reusing information related to data processing, consent, legal bases, data subjects, and rights. Using the DPV aids in mapping and describing legal obligations, such as those outlined in the GDPR. By reusing the DPV vocabulary, we ensure a consistent interpretation of requirements such as consent management, lawful processing, and data subject rights. This vocabulary can be integrated into software tools to automate privacy compliance tasks, including auditing, reporting, and managing privacy settings in accordance with legal frameworks. Furthermore, the DPV can be expanded to meet specific domain requirements, allowing organisations to address custom privacy needs or sector-specific regulations.

Overall, the Data Privacy Vocabulary enhances legal compliance, improves transparency, facilitates automation, and ensures consistent handling of privacy-related data across various systems and jurisdictions. The DPV also represents a community consensus regarding the modelling of concepts and their semantic expression through the DPVCG.

The manual process of matching GDPR concepts to Data Privacy Vocabulary (DPV) terms aimed to identify exact, narrower, or no matches. This process drew on François Scharffe's 2009 work on ontology alignment [24] . The glossary of terms identified in Table 19 (gathered from the Regulator templates (see Section 4.5.2) was quite helpful here as it showed a high level of consistency between regulator templates, with many terms being consistent across templates (for example, eighteen terms appear in all templates). For the mapping of GDPR concepts, each GDPR concept was compared with existing DPV concepts to see if there was an exact match, a partial (narrower) match or no match. The matching process considered the GDPR article number (if quoted) and the context of use to support the matching. If the concept was missing from the DPV this was categorised as no match. A sample of the matching process for GDPR concepts to DPV is displayed in Table 22.

Table 22 Example of GDPR Concept Matching Process with DPV Ontology.

GDPR Concept	DPV term	Match Status
Legal Basis	dpv:LegalBasis	Exact Match
Name of Business Process	dpv:PersonalDataHandling	Partial Match
Data Breach	Proposed: ropa:DataBreachRecord	No match – propose as new term to DPVCG

There were three cycles conducted for matching of GDPR concepts to DPV to complete the CSM-RoPA , all of which were completed by the researcher. The first cycle, completed in 2020, consisted of 6 RoPA templates and identified 43 GDPR concepts. The second cycle, completed in 2022, consisted of seventeen templates, and identified 47 GDPR concepts. For the second cycle the original six templates were reviewed (for changes and updates) and eleven additional templates were identified and reviewed. The third matching cycle, completed in 2024, reviewed the same seventeen templates (for changes and updates), and identified 47 GDPR concepts. Each of these matching cycles is discussed below.

The findings of the first mapping cycle in 2020 are presented in Table 23. This cycle identified fourteen unique fields in GDPR which had exact matches with DPV, fifteen had partial matches, three had complex partial matches, whereas eleven unique fields had no match with DPV (version 0.1). Based on the mapping outcome, it was identified that eleven additional GDPR concepts should be proposed to be added to the DPV to map these ROPA templates. Among the 11 additional concepts required are International Transfers, Controller Name and Contact Details, Original Source of Data, Data Protection Officer, Data Protection Impact Assessment, Data Subject Rights, Risk, Privacy Notice, Representative & Data Breach Record, refer to online resource[34] for a complete list of the concepts mappings and relevant regulator templates. The output of this process was the creation of CSM-RoPA version 0.1.

Table 23 First Mapping GDPR concepts to DPV terms (2020).

GDPR Regulation	Combined ROPA Model Field	Mandatory Article 30 GDPR	Related DPV Concept	DPV mapping outcome
30	Register of Processing Activities	Y	No DPV Concept	None
30(1)(a)	DataController	Y	dpv:DataController	Exact
30(1)(a)	Controller name and contact details	Y	Many suitable vocabularies	None
30(1)(a)	Data Protection Officer	Y	No DPV Concept	None
30(1)(a)	Representative	Y	No DPV Concept	None
30(1)(a)	Joint Controller	Y	dpv:DataController	Partial

[34] https://doi.org/10.5281/zenodo.14914848

GDPR Regulation	Combined ROPA Model Field	Mandatory Article 30 GDPR	Related DPV Concept	DPV mapping outcome
30.1	Business Process	N	dpv:PersonalDataHandling	Partial
30.1	Owner of Process	N	dpv:DataController	Partial
30.1(b)	Purposes of processing	Y	dpv: Purpose	Exact
6.1	Legal Basis for Processing	N	dpv:LegalBasis	Exact
30 (a)	Type of Processing	N	dpv:Processing	Exact
30.1(c)	Categories of personal data	Y	dpv:PersonalDataCategory	Partial
9.1	Special Category Personal Data	N	dpv:SpecialCategoryPersonalData	Partial
30.1(c)	Categories of data subjects	Y	dpv:DataSubject	Exact
9.1	Vulnerable Data Subject Category	N	dpv:DataSubject	Partial
-	Classification Level	N	dpv:TechnicalOrganisationalMeasure	Partial
30.1(f)	Retention/Deletion Periods	Y	dpv:StorageDuration, dpv:StorageDeletion	Exact
6/14/30.1(b)	Data Combination	N	dpv:Combine	Exact
5.1	The source of data	N	No DPV Concept	None
28	Processor	N	dpv:DataProcessor	Exact
28.3	Data Processing Agreement	N	dpv:Contract	Partial
	Data Transfer	N	dpv:Transfer	Exact
28/30.1(c)	Data Categories subject to transfer	N	dpv:PersonalDataHandling, dpv:Transfer, dpv:PersonalDataCategory	Complex, Partial
30.1(d)	Categories of recipients of transfer data	Y	dpv:Recipient	Exact
30.1(e)	Third countries that personal data are transferred to	Y	dpv:location	Complex, Partial
44-47	Nature of Transfer to Third Country	N	dpv:LegalBasis	Partial
30.1(e)	Appropriate Safeguards for Third Country Transfers,	Y	dpv:TechnicalOrganisationalMeasure	Partial
32	Technology Used	N	dpv:TechnicalOrganisationalMeasure	Partial
35	Risk and Mitigation Measures	N	dpv:TechnicalOrganisationalMeasure, dpv:RiskManagementProcedure.	Complex, Partial
35	Risk - Information about the risk	N	No DPV Concept	None
30.1(g)	Technical and organisational measures of security	Y	dpv:TechnicalOrganisationalMeasure	Exact
35	Data Protection Impact Assessment	N	No DPV Concept	None
13/14/15	Data Subject Rights	N	No DPV Concept	None
13	Privacy Notice	N	No DPV Concept	None
6.1(f)	Legitimate interests for the processing	N	dpv:LegalBasis	Partial
6.1(f)	Legitimate Interest Assessment	N	dpv:LegalBasis	Partial
22.1	Automated decision-making	N	dpv:Processing	Exact
6.1	Link to the record of consent	N	dpv:consent	Exact
5	Location of personal data	N	dpv:StorageLocation	Exact
30.1	Status of processing	N	dpv:PersonalDataHandling	Partial
33.5	Personal Data Breach	N	No DPV Concept	None
30.1(f)	Retention and erasure policy.	N	TechnicalOrganisationalMeasure StorageRestriction	Exact
36.1	Prior Consultation with DPA	N	No DPV Concept	None

GDPR Regulation	Combined ROPA Model Field	Mandatory Article 30 GDPR	Related DPV Concept	DPV mapping outcome
30.1(b)	Main or Auxiliary Processing activity	N	Purpose	Partial

A second mapping cycle was conducted in 2022 to create a wider and more comprehensive model of RoPA. This cycle reviewed thirty-one regulator websites and identified 17 RoPA templates. The mapping of concepts to the DPV terms used the same process as the first cycle. A summarised outcome of the 2022 mapping is presented in Table 24, which found that forty-four concepts had exact matches, one concept was partially matched, and two concepts could not be matched with the DPV. The three concepts which were not exact matches were proposed to be added to the DPV. This was a significant improvement on the 2020 mapping as it covered additional regulator templates, and the previously proposed concepts had been added to the DPV, thereby increasing its coverage of RoPA information. The output of this process was the creation of CSM-RoPA version 0.2. Please refer to the online resource for the full detailed mapping of CSM-RoPA 0.2[35].

Table 24 Second Mapping of GDPR concepts to DPV terms (2022).

Concept defined in GDPR text	No of GDPR concepts	Example
Defined	44	'Legal basis'
Partially defined	1	'Name of Business Process'
Not defined	2	'Status of processing' and 'Data Breach Record'

The third mapping cycle was conducted in 2024. This cycle involved seventeen regulator templates (the same as the second cycle) with 47 GDPR concepts. The purpose of the 2024 mapping was to check whether the earlier partial and no matches were still appropriate, and to establish that full matching was achieved. This mapping cycle found a complete mapping of GDPR concepts with DPV concepts through forty-seven exact matches. The output of this process was the CSM RoPA version 0.3. The full mapping is provided in Appendix B.

In total, the three matching cycles between 2020 and 2024 have added twenty-six terms to the DPV and enabled the creation of CSM-RoPA to express GDPR concepts required for representing RoPA information. Table 25 presents a full summary of the three matching cycles.

[35] https://doi.org/10.5281/zenodo.14914848

Table 25 Summary of Outcomes of Mapping Cycles between GDPR Concepts and DPV

CSM-RoPA Version	Year	No. of Regulator Templates reviewed	No. of GDPR Concepts Identified	Extent of Coverage of GDPR concepts using DPV				Concepts contributed to DPV
				Exact Match DPV	Complex/Partial	Partial	No Match	
0.1	2020	6	43	14	3	15	11	26
0.2	2022	17	47	44	0	1	2	2
0.3	2024	17	47	47	0	0	0	-

The last mapping that resulted in the creation of CSM RoPA v0.3 is displayed in Table 26. The table contains the following information:

- The column 'GDPR' specifies the relevant clause in the GDPR,
- The Column '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 CSM-RoPA 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 could it be directly linked to the Accountability Principle, but was suggested to be used by the regulator by including it in the template.

Table 26 Mapping of CSM-RoPA 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 the 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	*E*	*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	*E*	*R*	*R*
35	Risk management	dpv:RiskMitigationMeasure	*E*	*R*	*O*
35	Relevant DPIA	dpv:DPIA	*E*	*R*	*R*
35	DPIA Results	dpv:DPIA	*E*	*R*	*O*
36.1	Impact Assessments	dpv:ImpactAssessment	*E*	*R*	*R*
36.1	Prior Consultations	dpv:Consultation	*E*	*R*	*R*
37.6	External DPO organisation	dpv:DataProtectionOfficer	*E*	*R*	*R*
_	Name of Business Process	dpv:Process	*E*	*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/Auxiliary 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*

GDPR	Field	DPV	Map.	DC	DP
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*

Six additional competency questions have emerged as part of the review of regulator templates and ontology engineering. This brings to a total of fifty competency questions for CSM-RoPA. These questions are presented in Table 27 below.

Table 27 Competence Questions from Ontology Engineering Requirements

Question No.	Competence Question
CQ45	What information concepts are mandatory entries for RoPA?
CQ46	What are the Specific RoPA requirements for each regulator?
CQ47	What is the source of the information required for RoPA completion?
CQ48	Is there a requirement for regulators in different jurisdictions to have different RoPA outputs?
CQ49	How is the metadata for CSM-RoPA maintained?

5.2.5 Ontology Design

The CSM-RoPA ontology is designed to use a bottom-up approach [23]. In this approach, specific data, concepts, and instances in a domain are gathered and worked towards more generalised and abstract ideas rather than beginning with a predefined structure or high-level concepts. This method is grounded in the actual data or instances of knowledge, focusing on building the ontology based on the concrete information at hand instead of imposing a top-down, predefined hierarchy or schema.

For CSM-RoPA, general concepts are modelled using OWL as classes, and specific variations are subclasses. These subclasses inherit from their parent class, and all instances of a subclass must be instances of the parent class. The logical relationships in the domain and use are based on the level of granularity needed, domain-specific constraints, and reusability considerations. An example of this would be Special category personal data which would be a subclass of personal data. This provides for a clear, consistent, and logically structured ontology that accurately represents the knowledge in your domain. Developing classes, subclasses, and relationships is supported using

Protégé [179], a widely utilised visual editor for creating OWL ontologies. Figure 15 shows a relationship between class subclass and object properties for personal data processing [6] . This shows consent and legitimate interest as subclasses of Legal basis. Legitimate interest assessment is a subclass of Legitimate Interest.

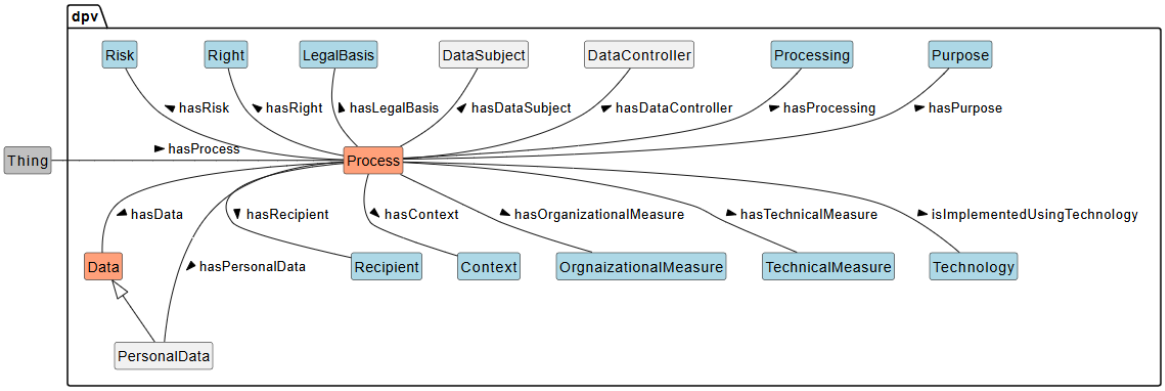

Figure 15 Classes and Subclasses to Represent Processing from DPV [6].

The first version of CSM-RoPA 0.1 was created in 2020 [13]. It incorporated six regulator templates, which provided 43 GDPR concepts that were modelled as twenty classes and seven properties. Of its twenty classes, all were reused from DPV. For GDPR concepts where no match was found with the DPV, these 11 GDPR terms were submitted to the DPVCG to improve the DPV's expressivity and representation of RoPA templates. These terms were incorporated into the next version of the DPV [6]. The eleven additional concepts required were International Transfers, Controller Name and Contact Details, Original Source of Data, Data Protection Officer, Data Protection Impact Assessment, Data Subject Rights, Risk, Privacy Notice, Representative & Data Breach. CSM-RoPA version 0.2 was created in 2022 [18]. It used seventeen regulator templates and contained forty-three classes and twenty-six properties. Of its forty-three classes, forty-one were reused from DPV, and two newly created concepts were submitted to the DPVCG and incorporated into the next version of the DPV. The concepts were data breach and the name of the business process.

The current/latest version, CSM-RoPA 0.3, was created in 2024 and used seventeen templates to create forty-five classes and twenty-six properties. Of its forty-five classes, forty-four were reused from DPV and one from Dublin Core standard[36]. CSM-RoPA 0.3 provides a complete mapping and coverage of concepts to represent the information found in RoPA.

Table 28 displays the three cycles of mapping and shows the progression of the mapping over the three cycles. The column titled GDPR refers to the GDPR article. The column called field refers to the GDPR concepts identified on regulator templates. The Mapping cycle refers to the three cycles for creating CSM-RoPA: 0.1 in 2020, 0.2 in 2022, and 0.3 in 2024. In the mapping

[36] The class dct:contactPoint was reused from the Dublin Core Vocabulary

columns, *E* means exact match, *Nil* means identified but no match, *Pt* means a partial match, *CX*, *Pt* means a complex partial match, and *X* means that the term was not identified in the 2020 mapping. Please refer to the online resource for the entire table with all terms from each mapping cycle online.[37]

Table 28 Overview of Evolution of CSM-RoPA Mapping Outcomes.

GDPR	**Field**	**Mapping cycle**			**Comment**
		2020	2022	2024	
5	Location of personal data	E	E	E	2020 - 2024 mapping consistent
5.1	Data Sources	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
6.1	Legal basis	E	E	E	2020 - 2024 mapping consistent
6.1	Link to the record of consent	E	E	E	2020 - 2024 mapping consistent
9.1	Special Personal Data	Pt	E	E	Consistency of term agreed with DPVCG
9.1	Vulnerable Data Subjects	Pt	E	E	The term was added to DPV following the 2020 mapping
22.1	Automated decision-making or profiling	E	E	E	2020 - 2024 mapping consistent
26.1	Joint Controller agreement	X	E	E	First Identified and mapped in 2022
28	Data Processors	E	E	E	2020 - 2024 mapping consistent
28.3	Data Processing Contract	Pt	E	E	The term was added to DPV following the 2020 mapping
28.3	Data processor contract	X	E	E	First Identified and mapped in 2022
30.1	Status of processing	Pt	S	E	The term was added to DPV following the 2020 mapping
32	Tech/Org measures implementation	X	E	E	First Identified and mapped in 2022
32	Security measures	X	E	E	First Identified and mapped in 2022
32	Technologies used	Pt	E	E	The term was added to DPV following the 2020 mapping
33.5	Data Breach	Nil	S	E	The term added to DPV for the 2024 mapping
35	Risk management	Cx, Pt	E	E	The term was added to DPV following the 2020 mapping
35	Relevant DPIA	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle

[37] https://doi.org/10.5281/zenodo.14914848

GDPR	Field	Mapping cycle			Comment
		2020	2022	2024	
35	DPIA Results	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
36.1	Impact Assessments	X	E	E	First Identified and mapped in 2022
36.1	Prior Consultations	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
37.6	External DPO organisation	X	E	E	First Identified and mapped in 2022
_	Name of Business Process	Pt	Pt	E	The term added to DPV for the 2024 mapping
_	Owner of Process	Pt	E	E	Utilised Dublin Core vocabulary
_	Type of Processing	E	E	E	2020 - 2024 mapping consistent
13, 14, 15	Data Subject Rights	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
28, 30.1(c)	Data Categories Transfer to Third Parties	Cx, Pt	E	E	The term was added to DPV following the 2020 mapping
30(1)(a)	DataController	E	E	E	2020 - 2024 mapping consistent
30.1(a)	DPO contact	X	E	E	First Identified and mapped in 2022
30.1(a)	Representative	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
30.1(a)	Representative Contact	X	E	E	First Identified and mapped in 2022
30.1(a)	Name of joint controller	Pt	E	E	The term was added to DPV following the 2020 mapping
30.1(a)	Contact details of joint controller	X	E	E	First Identified and mapped in 2022
30.1(b)	Purposes of processing	E	E	E	2020 - 2024 mapping consistent
30.1(b)	Main/Auxiliary Processing	Pt	E	E	The term was added to DPV following the 2020 mapping
30.1(c)	Categories of data subjects	E	E	E	2020 - 2024 mapping consistent
30.1(d)	Categories of Recipients	E	E	E	2020 - 2024 mapping consistent
30.1(c)	Personal Data Categories	Pt	E	E	Consistency of term agreed with DPVCG
30.1(f)	Retention/Deletion Periods	E	E	E	2020 - 2024 mapping consistent
30.1(e)	Third Countries Data Transfer	Cx, Pt	E	E	The term was added to DPV following the 2020 mapping
30.1(g)	Technical and organisational measures	E	E	E	2020 - 2024 mapping consistent
30.1(e)	Appropriate Safeguards	Pt	E	E	The term was added to DPV following the 2020 mapping
30(1)(a)	Data Controller Contact	Nil	E	E	Not represented in 2020, added in 2022

GDPR	Field	Mapping cycle			Comment
		2020	2022	2024	
30(1)(a)	Data Protection Officer	Nil	E	E	Not present on 2020 templates - mapped in the 2022 cycle
44–47	Nature of Transfer	Pt	E	E	The term was added to DPV following the 2020 mapping
6.1(f)	Legitimate interests	Pt	E	E	The term was added to DPV following the 2020 mapping
6.1(f)	Legitimate interests assessment	Pt	E	E	The term was added to DPV following the 2020 mapping
6, 14, 30.1(b)	Data Combination	E	E	E	2020 - 2024 mapping consistent
30	Record of Processing Activities	Nil	E	E	The term was added to DPV following the 2020 mapping

The exercise of creating the CSM-RoPA led to improving the coverage of DPV to represent RoPA concepts. An example of this is the GDPR concept of 'data breach record', for which DPV in 2020 had no matching concept. This term was created from the work presented in this thesis as dpcat:DataBreachRecord, which was then proposed and added to the DPV. Finally, in 2024, the term was added to the DPV version 2.0 as *dpv:DataBreachRecord*. Another example is the GDPR concept of 'Third country data transfers,' a complex partial match with DPV in 2020 based on using a combination of country and *dpv:Location.* The concept *dpv:ThirdCountry* was proposed and accepted to enable an exact representation of the concept as required in RoPA and GDPR. The full 2024 mapping representation and relationships are presented in Appendix C, and a complete graphical representation is provided in Figure 16.

Table 29 Sample of CSM-RoPA 0.3 Classes and Relationships 2024

(see Appendix C for the full table)

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
Data Sources	dpv:DataSource	The source or origin of the data	dpv:hasDataSource	dpv:PersonalDataHandling	dpv:DataSource	R
Legal Basis	dpv:LegalBasis	Legal basis used to justify processing of data or use of technology by a law	dpv:hasLegalBasis	dpv:PersonalDataHandling	dpv:LegalBasis	M
Data Processor	dpv:DataProcessor	A 'processor' means a natural or legal person, public authority, agency, or other body which processes data on behalf of the controller.	dpv:hasDataProcessor	dpv:PersonalDataHandling	dpv:LegalEntity	M
Technical/Organisational measures	dpv:Technology	The technology, technological implementation, or any techniques, skills, methods, and processes used or applied	dpv:hasTechnicalOrganisational Measure	dpv:PersonalDataHandling	dpv:TechnicalOrganistional Measure	M
Data Categories Transfer to third parties	dpv:Transfer, dpv:PersonalData	The personal data categories that are transferred to third-party recipients	dpv:hasRecipient	dpv:PersonalDataHandling	dpv:Transfer	R

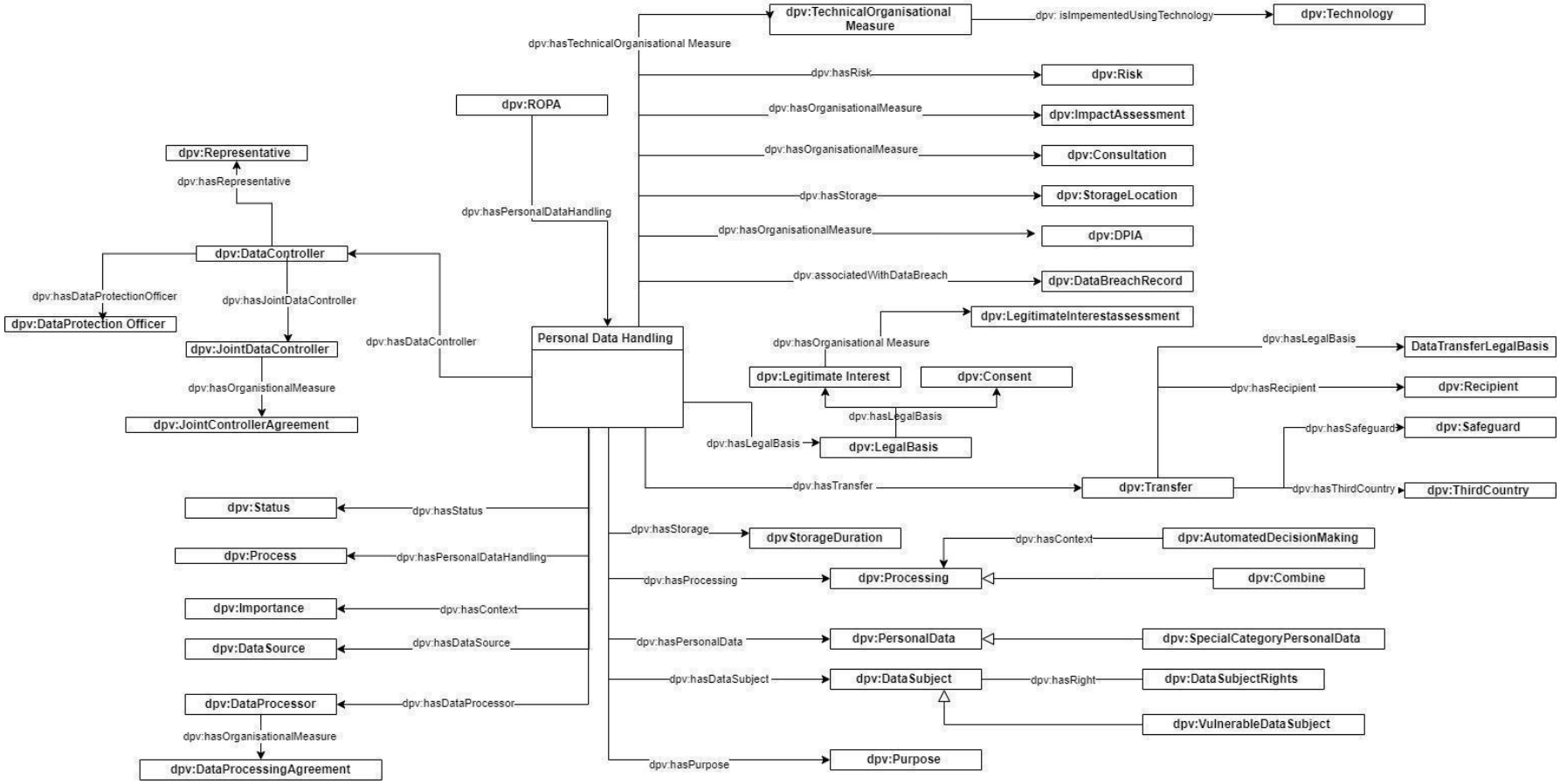

Figure 16 A Graphical Representation of CSM-RoPA 0.3.

5.3 Ontology Implementation

This section describes converting the mapped GDPR concepts described in section 5.2 into an ontology language to develop the CSM RoPA ontology. The process of generating the CSM-RoPA used the Protégé ontology editor to create the classes and properties and document them with human-readable annotations. The classes were represented as instances of rdfs:Class, owl:Class and skos:Concept, with a similar approach taken for properties. Each class and property had labels added as per WIDOCO best practices. The terms missing from the DPV were then proposed to the DPVCG to be added to the DPV and were subsequently integrated into the next versions of the DPV. Examples of two terms added to the DPV are presented in Appendix D, which presents a turtle representation of a data protection officer. Appendix E provides a serialised RDF representation of the object property 'hasDataProtection Officer.' Both examples also contain the provenance maintained by the DPVCG regarding the creation of the terms which reflects the author of this thesis as the contributor for those terms.

The CSM-RoPA ontology is documented using WIDOCO, which is an ontology documentation tool that allows users to publish and create customised documentation for an ontology [27]. Given an ontology, WIDOCO detects metadata and creates documentation with diagrams, human-readable descriptions of the ontology terms, and a summary of changes relative to previous versions of the ontology. The documentation consists of linked enriched HTML pages that end users can further extend. WIDOCO is open source and builds on well-established Semantic Web tools. The CSM-RoPA 0.3 ontology and documentation generated through this process can be found here.

5.4 Evaluation

This section evaluates the CSM-RoPA model, and seeks to establish the following:

- Does the ontology contain the necessary concepts and relationships to represent a GDPR RoPA?
- Does the CSM-RoPA ontology meet Semantic Web standards and best practices?

The evaluation consists of five parts, which are listed below:

1. An analysis of the extent to which CSM-RoPA ontology meets the competence questions set out in the ontology specification (see Chapter 5.4.1).
2. A review of the extent to which the CSM-RoPA ontology meets Semantic Web best practices.
3. A use case demonstrating the extent to which CSM-RoPA can express a RoPA.

4. An overview of the presentation of CSM-RoPA to the DPVCG peer review
5. An overview of the peer-reviewed publications and industry use of CSM-RoPA

5.4.1 Answering the Competency Questions

The first section of the evaluation establishes the extent to which CSM-RoPA meets the competency questions the ontology must meet as set out in the specification (see Section 5.2.2). Table 30 provides an example of how the competency questions CQ1-CQ44 can be answered by enhancing the DPV with concepts identified through the CSM RoPA process (refer to Appendix F for the full table of CQ1-CQ44). The CSM-RoPA thus meets all the competency questions required in the specification.

Table 30 Meeting the Competence Questions using terms from CSM-RoPA

Question No.	Competency Questions	CSM-RoPA Class	CSM-RoPA Property
CQ01	What is the purpose of the data processing activity?	dpv:Purpose	dpv:hasPurpose
CQ02	What categories of personal data are being processed?	dpv:PersonalDataCategory	dpv:hasPersonalData
CQ03	Who is the data controller responsible for the processing activity?	dpv:DataController	dpv:hasDataController
CQ04	Who are the data processors involved in the processing activity?	dpv:DataProcessor	dpv:hasDataProcessor

CSM-RoPA can fully meet competence questions CQ1-CQ44, which comes from GDPR article 30 and from regulator template requirements. The ontology engineering for CSM-RoPA identified five further competence questions (see Section 5.2.4). These are CQ45-CQ49 and are addressed below.

CQ45 What information concepts are mandatory entries for RoPA? The mapping of GDPR RoPA Concepts to DPV terms (2024) provided in Table 29 summarises the mapping between CSM-ROPA fields and DPV concepts. The mapping table represents the necessity of field each for Data Controllers and Data Processors RoPA

CQ46 What are the specific RoPA requirements for each regulator? The legal responsibility for organisations based on GDPR Article 30 requires that only the mandatory requirements as set out in the article are met. However, guidance from regulators advises that RoPA be supplemented with the necessary information to demonstrate accountability. Hence, a template provided by a regulator could be considered the minimum threshold required to demonstrate accountability for that regulator.

CQ47 What is the source of the information required for RoPA completion? The information necessary to populate the RoPA resides within GDPR accountability documents. These documents include privacy notices, data processing agreements and GDPR policy documents [180]. Completing

RoPA requires the collection and representation of this data and transfer to RoPA. (This is discussed in Chapter 6).

CQ48 Is there a requirement for regulators in different jurisdictions to have different RoPA outputs? The use case in 6.4.3 provides an example of a RoPA output using CSM-RoPA, using SPARQL query. This could be extended as required to add additional RoPA fields to meet individual regulatory template requirements.

CQ49 How is the metadata for CSM-RoPA maintained? CQ47 discusses the sources of RoPA information. Organisations are best served to maintain data management capability [41] and a data catalogue approach (see Section 3.2) to maintain the metadata required for CSM-RoPA.

5.4.2 Following Ontology Engineering Best Practices

The second section of the evaluation discusses the best practices for developing the CSM-RoPA ontology. The ontology was developed using best practices for publishing linked data [181] .The CSM-RoPA ontology is not used directly by the end user, but its outcomes are used via the DPV. CSM-RoPA is modelling concepts that are intended to be submitted to the DPVCG. Each new concept that has been identified has been submitted to the DPV. Hence, the CSM-RoPA ontology does not need to be published separately.

Ontology Documentation: The CSM-RoPA ontology used the WIDOCO wizard for documenting ontologies. The creation of the ontology followed WIDOCO's best practice [27]. The wizard was used to detect missing vocabulary metadata and create documentation with diagrams, human-readable descriptions of the ontology terms and a summary of changes for previous ontology versions. The documentation consists of a set of linked enriched HTML pages that end users can further extend. WIDOCO is open source and builds on well-established Semantic Web tools.

Best practice for ontology publication: The FAIR 'Findable, Accessible, Interoperable and Reusable' principles provide guidelines and best practices for describing and accessing research data to be reused by others[28]. For the evaluation of CSM-RoPA, the FOOPS ontology pitfall scanner [147] for FAIR is utilised to validate best practices for publication of ontologies on the web. The validator provides the means for researchers to assess whether a vocabulary (OWL or SKOS) conforms to the best practices for publishing ontologies on the Web.

CSM-RoPA follows best practices and guidelines established within the semantic web community. CSM-RoPA gains best practice benefits using concepts from DPV. Some of these benefits are W3C Best Practices for Publishing Linked Data using WIDOCO [27] best practices with FOOPS for evaluation [147], W3ID for permanent IRIs and GitHub for version control and collaboration that comes with DPV. An evaluation of CSM-RoPA concepts using FOOPS showed

CSM-RoPA scored like the DPV, indicating consistent standards with the DPV. The benefits of reusing concepts from DPV are that concepts must adhere to a standard documentation approach for inclusion. This DPV standard sets out terms, labels, descriptions, comments, and statuses for each concepts seeking inclusion in DPV[38]. These concepts are reviewed and approved by the DPVCG before acceptance to DPV, thus maintaining the integrity and quality of the DPV. This approach ensures that CSM- RoPA concepts are well-documented and defined.

5.4.3 Use Case – Representing a RoPA

In this section, a use-case is conducted to take a RoPA document (typically presented as a spreadsheet based on the researcher's experience as a practising DPO) and retrieve information from it. The information in the RoPA was represented in RDF using CSM-ROPA, put in a triple store, and then retrieved using SPARQL queries based on CSM-ROPA concepts. This task was completed and presented in this section displayed.

For the use case, a sample RoPA is created to reflect the practical use of CSM-ROPA in an organisation. The use case uses prepared test data based on established organisational RoPA practices and processing activities. A copy of all data sets is available on GitHub [39]. The prepared data includes multiple scenarios, such as various organisational units conducting various organisational processing activities using external processors to reflect reality. It involves the following steps.

1. Generate test RoPA record data in standard English language (manually)
2. Convert data to RDF as a machine-readable format by using CSM-ROPA (manually)
3. Import to GraphDB[40] knowledge graph.
4. Generate RoPA from the knowledge graph using a SPARQL query.
5. Review output contains all information within the RoPA record

The use case successfully expresses manual English language RoPA records, loads the records to acknowledge a knowledge graph using CSM-RoPA and generates a RoPA from the knowledge graph. The processing records are converted to RDF in Turtle format and loaded to the triple store to create a knowledge graph, using the import function of GraphDB. Once the data is loaded into the knowledge graph, SPARQL queries are used to create 'views' for presenting a summary and overview of activities of information required for ROPA, their temporal periods, and the contact point for further communication. The SPARQL query is available on GitHub[41] and In Appendix G. The sample RoPA generated is displayed in Table 31.

[38] https://github.com/w3c/dpv/blob/master/code/style-guide.md
[39] https://github.com/coolharsh55/DPCat/tree/master/SEMANTiCS2021-DPCatalogue/sample-datasets
[40] https://graphdb.ontotext.com/
[41] https://github.com/coolharsh55/DPCat/blob/master/SEMANTiCS2021-DPCatalogue/ropa-view.sparql

Table 31 Sample Extract of Controller ROPA.

Department	**Customer Service Dept.**	**HR Dept.**	**Marketing Dept.**
Title	Record001	Record004	Record001
Period Start	2019-01-01	2019-01-01	2019-01-01
Period End	2022-12-13	2022-12-13	2022-12-13
Contact Name	Alice	Bob	Emily
Contact e-mail	alice@example.com	bob@example.com	emily@example.com
Purpose Category	Customer care	Service Provision	Direct Marketing
Purpose	Recording of customer calls	Expenses activities	Direct marketing via e-mail
Data Subject	Customers	Employees	Customers
Personal Data Category	Voice recordings	Financial	E-mail addresses
Recipient	Null	Beta Ltd.	Null
Recipient Category	Null	Data Processor	Null
Recipient Location	Null	Canada	Null
Storage years	2.0	7.0	1.0
Measures	Standard	Standard	Standard

This demonstrates that CSM-RoPA includes the necessary information concepts, relationships, and axioms to represent a GDPR RoPA. The objective was fulfilled by representing the manually generated ROPA information in a machine-readable format.

5.4.4 Peer Review in W3C DPVCG

In February 2021, CSM-RoPA 0.1 was presented to the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), comprising technologists and legal experts. The presentation provided an overview of the development of CSM-RoPA, an analysis of how CSM-RoPA can meet the requirements of the ICO accountability framework, and a summary of CSM-RoPA publications up to that point. The DPVCG offered feedback and peer review on terms that needed to be added to DPV and suggested that further work should be done to gather all available templates for inclusion in the analysis. In 2022, the updated mapping for CSM-RoPA 2022 was presented to the DPVCG, and two new concepts were introduced to the DPVCG for approval. In total, twenty-six new concepts were added to the DPV (see table 26). Each of these concepts required appropriate documentation to be submitted to DPVCG, and each concept was discussed and analysed by the group before being added to the DPV.

5.4.5 Peer-Reviewed Publications and Industry Uptake

The findings from this research have been presented in three peer-reviewed publications and conference presentations (see Table 32).

Table 32 Peer-Reviewed Publications concerning the Semantic Model of RoPA.

Publication	**Author Type**	**Venue**	**Citations (Feb 2025)**
A Common Semantic Model of the GDPR Record of Processing Activities [13]	Lead	Legal Knowledge and Information Systems: JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9-11, 2020	9
Demonstrating GDPR Accountability with CSM-RoPA: extensions to the data privacy vocabulary [33]	Lead	24th International Conference Enterprise Information Systems (ICEIS '21), 26-28 Apr 2021	4
Support for enhanced GDPR accountability with the standard semantic model for RoPA (CSM-RoPA)[19]	Lead	Springer Nature Computer Science Journal 3 (3), 224	11
Data Privacy Vocabulary (DPV) - Version 2 [6]	Co-author	The 23rd International Semantic Web Conference (ISWC 2024)	21

The GDPR RoPA concepts generated through CSM-RoPA and added to the DPV are now used in academic and industrial projects (see Section 1.7).

- Signatu develops automated solutions for GDPR compliance using the RoPA concepts widely [36].
- Alias Law, the developers of URoPA[42], have expressed an interest in reusing this work for further research.
- This work has contributed to a new ISO standard for RoPA under development.

5.5 Learning from the Development and Evaluation of the Semantic Model

The reflection and learning stages are essential in the Action Design Research (ADR) methodology. This chapter concerns the development of an ontology to implement the CSM-RoPA using Semantic Web standards and best practices. The key learnings gathered are as follows:

[42] https://github.com/uropa-project/uropa

- The utilisation of concepts from the DPV has proven to be successful for the development of CSM-RoPA. Using DPV to express RoPA concepts ensures that best practices for ontology engineering are met.
- The CSM-RoPA ontology successfully represented the concepts required for RoPA. However, additional work is required to collect and transfer RoPA information from source documents and stakeholders to the RoPA (refer to Chapter 6).
- CSM-RoPA forms the basis for the representation but requires support to enable interoperability, verification, and validation of RoPA information (refer to Chapter 6).
- The metadata required for conversion of GDPR information in source documents requires the organisation to have a data management capability (refer to Chapter 7.2.3).

5.6 Conclusions

This chapter addressed RSQ2.b to develop an ontology for implementing the CSM-RoPA component of the ERoPA Approach based on semantic-web standards and best practices. The chapter used the NeOn methodology to specify and create the semantic model of RoPA to represent information concepts to represent GDPR processing activities. The model was evaluated to establish the extent to which the ontology contains the necessary concepts, relationships, and axioms to represent a GDPR RoPA and to show that CSM-RoPA meets Semantic Web standards and best practices. The evaluation establishes that the model successfully meets the specified competence questions, meets ontology best practices, and is demonstrated to represent RoPA. The ontology has been peer-reviewed, published, and presented to many stakeholders and has been adopted and used within industry.

6 Interoperable RoPA Specification

6.1 Chapter Overview

This chapter details the creation and assessment of the Data Processing Catalogue specification (DPCat) to enable the **interoperable representation and communication** of RoPA information between data protection stakeholders. The DPCat interoperability specification builds on the CSM-RoPA ontology (see Chapter 5). The role of the Interoperability specification within the ADR methodology is illustrated in Figure 17. This chapter covers the research sub question RSQ2.c: to develop a specification for the interoperable exchange of RoPA information between stakeholders based on user requirements and best practices. The chapter also demonstrates the application of DPCat to five common data protection activities to evaluate the extent to which the application of the DPCat specification supports the requirements for implementing GDPR accountability.

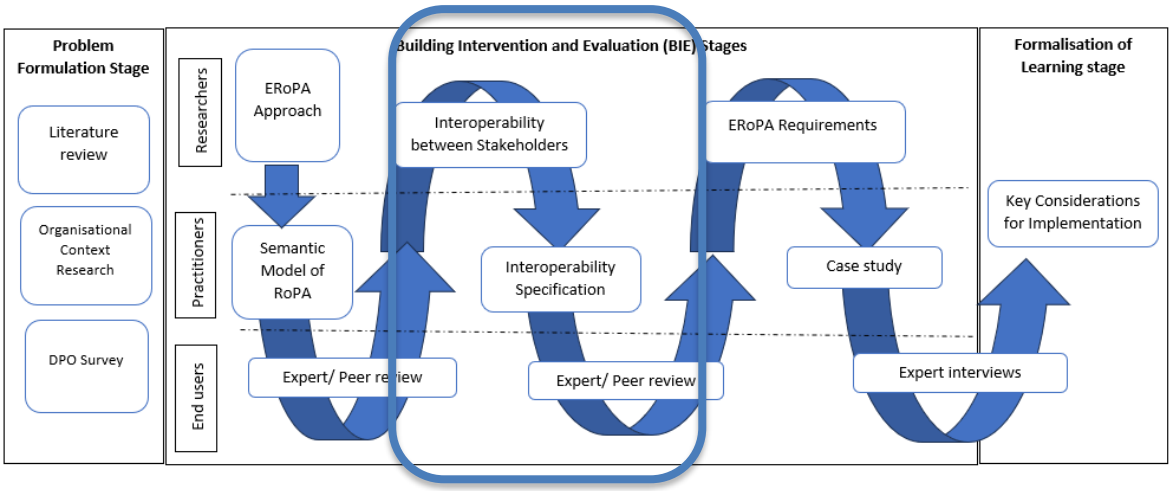

Figure 17 Second Build, Implement and Evaluate Stage of ERoPA Development

The chapter addresses the second ADR BIE stage (see Figure 17). Section 6.2 presents the use cases requiring an interoperable specification for RoPA. This is based on requirements gathered in Section 4.7. Section 6.3 provides the design for the Interoperable RoPA, and the ontology engineering methodology used, Section 6.4 presents its implementation, and Section 6.5 evaluates the specification through a case study implementation of representing and communicating a RoPA using DPCat for the completion of five common data protection tasks. Section 6.6 then formalises the ADR learning from the second BIE stage.

6.1.1 ADR Roles

The development of the DPCat specification utilises designated ADR Roles to support its development as a component of the ERoPA Approach (see Table 33). This first of these roles, the researcher, a practising DPO with experience in conducting the planning and implementation of the ERoPA, conceptualises the DPCat specification, while the second role, the Practitioner role, is met by the Data Privacy Vocabularies and Controls Community Group (DPVCG), who contribute their expertise and context to the development and implementation of the ERoPA artefact. The third ADR role, " users" for the DPCat specification, is formed of (i) peer-reviewed publications where experts provide their feedback, (ii) industry usage by organisations, and reference to the (iii) original DPO survey requirements.

Table 33 ADR Role Assignment for BIE 2

Role	Assignment
Researcher	The Thesis Researcher (a practising DPO)
Practitioner	Data Privacy Vocabularies and Controls Community Group (DPVCG)
User	Peer-reviewed publications, Industry usage, Original DPO survey requirements

6.2 Use Cases

The previous chapter enables the representation of a RoPA in a machine-readable and interoperable manner through the CSM RoPA ontology based on extending the DPV and covers information requirements from the GDPR and DPA RoPA templates. However, a RoPA is not a single document in practice but 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, or vendors - where such 'external entities' provide the information required for establishing records of agreed activities and assurance of compliance obligations. The ROPA provides the DPO with an essential overview of the organisation's practices and is part of the DPO's obligations regarding compliance (GDPR Art. 39) [11], [176]. 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.

Since the GDPR does not dictate or is concerned with how a ROPA is generated or maintained, the organisation can determine practices that suit its compliance approach and style to ensure it meets the legal requirements. For example, the organisation may maintain ROPAs centrally overseen by the DPO, where information from all sources is fetched and collated into a common document (e.g. a spreadsheet) and added to a single common information management system. Alternatively, the organisation may maintain separate RoPA documents for each department. In either case, upon being asked by a DPA or an auditor, the organisation must produce a RoPA for the specified criteria, such as reflecting specific processing activities or a certain period. To do this, the organisation must first identify and extract the relevant information from its RoPA documents. This task can involve manual efforts by the DPO or responsible entity unless the organisation uses technological solutions that support such use cases and provide an easier workflow based on automation.

The need to collect, represent and transfer RoPA thus motivates the need to create an interoperability specification to enable the flow of GDPR accountability data between RoPA stakeholders (see Section 4.7). The specification must enable gathering RoPA information from organisational units, third-party processors, and joint controllers to ensure the DPO has up-to-date and accurate information to complete their tasks. This also addresses the primary issue regarding RoPA in the regulator guidance and inspection reports (see Section 4.5.3).

This collected RoPA data must also be maintained in a format that can be shared with regulators and be available for inspection by regulators (see Section 4.5.3) as sharing RoPA with the auditors/certification bodies is becoming an emerging area of compliance (see Section 4.7). This specification should utilise an agreed-upon interoperability standard specification for the representation, collection and transfer of RoPA information in a machine-readable and interoperable manner to describe datasets so that publishers can use a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues (see Section 4.7 req. 2.2) [13], [18], [19], [33].

This section presents five use cases exploring key information flows, the involved stakeholders, and their roles regarding the ‘heterogeneous sources’ in RoPA-related data governance. This methodology follows prior work [33] [18] regarding identifying stakeholders and information flows related to GDPR compliance and establishing the utility of developing machine-readability and semantic interoperability mechanisms to facilitate communication and use of compliance information.

The obligations of GDPR article 39 consider 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; this section explores 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. The data controller's role is also considered 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 expressing ROPA-related information in a machine-readable and interoperable format.

Use Case U1: Data Controller

This use case, illustrated in Figure 18, 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. 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 exchanged between these stakeholders can involve: (i) A ROPA that conforms to GDPR Art.30 requirements; (ii) A ROPA that conforms to 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.

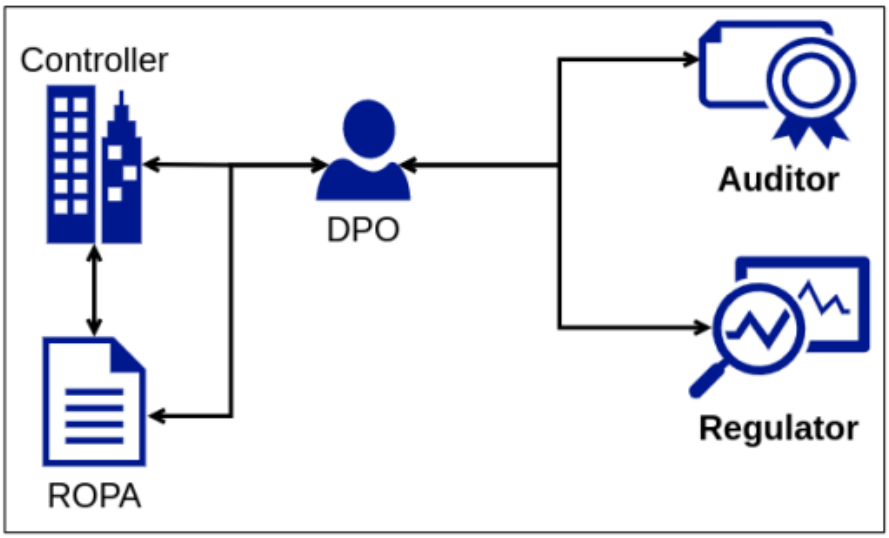

Figure 18 Basic generation of legal requirement ROPA [18].

Use Case U2 Data Controller with Internal Organisational Units

The second use-case, presented in Figure 19, expands U1 with internal information flows through two 'organisational units' or departments: Payroll and Marketing, where relevant data from each

unit must be collated to create a common organisation-level RoPA [18]. U2 also involves potential follow-ups with each unit regarding maintaining records per department and establishing 'points of contact' and 'responsible entities.' While the information flows to external entities, such as DPAs and auditors, and stays the same as U1, the internal organisational units are not separate legal entities and are not subject to direct investigation. The organisation is responsible for all internal units' RoPA obligations. The critical 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.

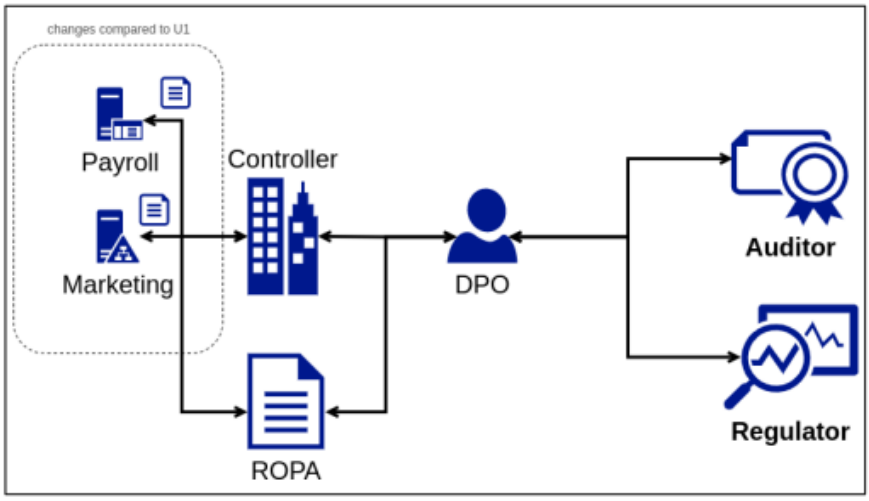

Figure 19 Organisational units updating and maintaining RoPA [18].

Use Case U3: Data Controller with Data Processors

The third use case, illustrated in Figure 20, has additional information flows where the controller and its DPO collect relevant information from appointed (external) processors. In cases where a processor is common to all departments or is managed at the organisational level, U3 is an extension of U1. However, in practice, different organisational units are likely to use specific external vendors (i.e. Data Processors), which makes U3 an extension of U2. In this, one can consider the practical situations where internal units often manage data governance 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, and contact details; (iii) Collation of information from heterogeneous sources into a common ROPA; and (iv) A selective part of ROPA, e.g. specific processor.

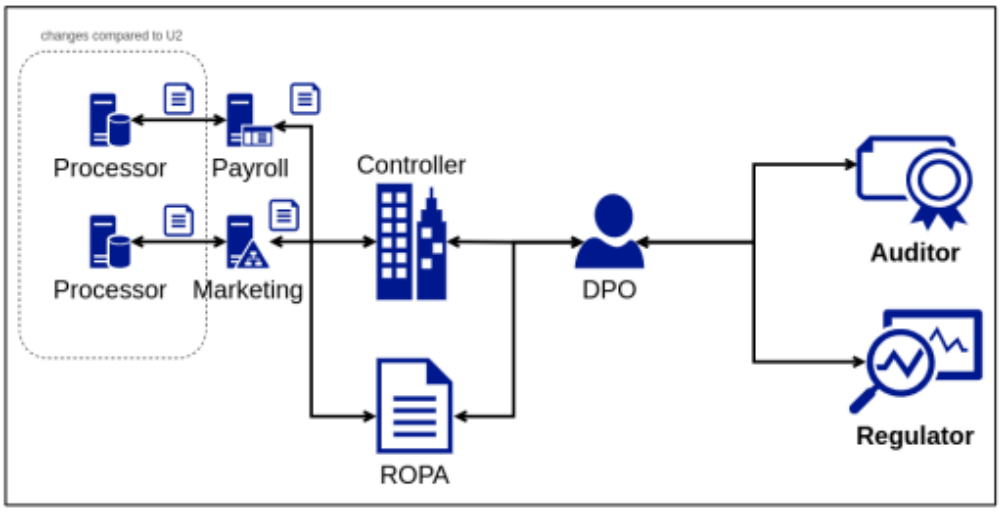

Figure 20 A data controller with organisational units and data processors [18].

Use Case U4: Data Controller in a Joint Controller Relationship

The fourth use case, illustrated in Figure 21, expands U3 with the Data Controller being in a Joint controller relationship with two or more controllers sharing the responsibility of processing as per GDPR Art. 26. Like the possibility of associating processors with organisational units in U3, joint controllers can also similarly be related to 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, and 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.

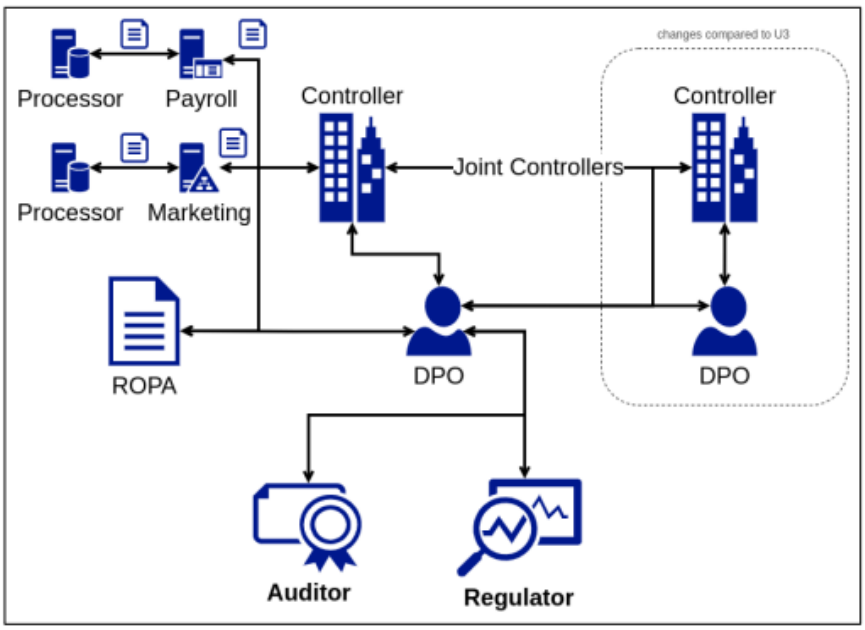

Figure 21 Data controller in a joint controller relationship [18].

Use Case U5: DPO Overseeing Multiple Data Controllers

Use cases U1-U4 considered the perspective of a Data Controller that employs a DPO to manage their ROPA information. In U5, illustrated in Figure 22, the scenario of a DPO being an external organisation or individual providing 'DPO-as-a-service' is presented. This research calls this entity 'External DPO' and considers their duties to involve overseeing multiple organisations[43]. The external DPO must 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, and 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.

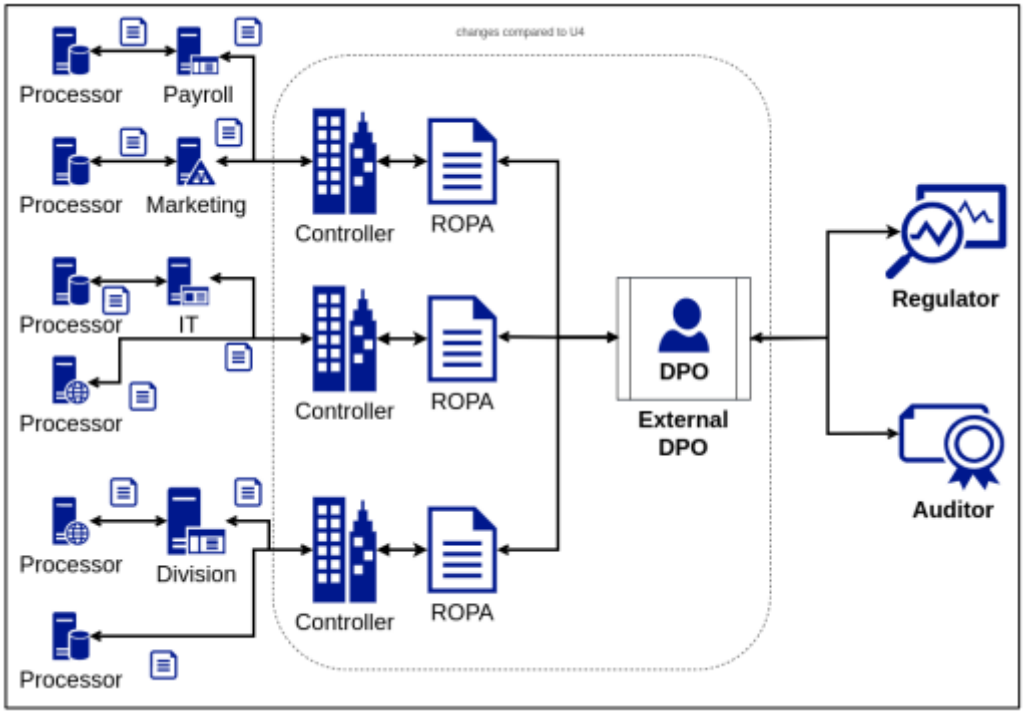

Figure 22 A DPO Overseeing Multiple Data Controllers [18].

In Chapter 5, the CSM-RoPA was created to represent the information required in a RoPA in a machine-readable and interoperable manner through Semantic Web ontologies. However,

[43] The author is an external DPO for several companies, and these use cases match the perceived work-related activities in typical organisations.

while CSM-ROPA is sufficient for generating a RoPA, additional information and actions are required for sharing and using its information amongst relevant stakeholders. The following section introduces the Data Processing Catalogue specification (DPCat) for gathering and sharing information as a 'catalogue' that can contain ROPA-related details and fulfil the requirements for its maintenance and exchange with stakeholders. This specification extends CSM-ROPA to enable collection, maintenance, and use of ROPA information from multiple stakeholders, produce ROPA with global and granular scopes for specific activities, services, or organisational units, enabling conformance and validation activities, and ensuring documentation of provenance such as sources, timestamps, and contact details.

6.3 DPCat Design

Based on the State of the Art (see Section 2.6) and the requirements for ERoPA (see Section 4.7), the Interoperability specification (DPCat) is crucial for the collection and transfer of RoPA and thus forms a vital component of ERoPA. DPCat builds on CSM-RoPA to provide a machine-readable conceptualisation of concepts essential for the exchange validation and conformance checking of RoPA information. It uses the NeOn methodology for the development of the specification [38] consisting of the following steps in Table 34.

Table 34 DPCat Specification Build Steps.

Steps	**NeOn Methodology Stage**	**Tasks**
1	Specify the requirements for the model - define the goals, scope, and requirements for the ontology.	Identify from analysis of State of the Art (see Section 4.7)
2	Knowledge acquisition - gathering of domain knowledge	Identify the GDPR concepts to be modelled from GDPR Article 30 (see Section 4.5.1) and regulator-supplied RoPA templates (see Section 4.5.2)
3	Ontology reuse and reengineering - reuse existing ontologies or ontology modules, modifying them to fit the current needs	Analyse existing ontologies (CSM RoPA, DPV) to represent RoPA information and standards such as DCAT to represent information collections.
4	Ontology design - develop the ontology, specifying classes, properties, instances, and axioms.	Create required classes and properties to represent the RoPA as a catalogue of information based.
5	Implementation - translate the ontology design into an actual implementation in an ontology language	Generate the ontology using RDFS and OWL and generate its documentation.
6	Evaluation - evaluate the ontology to ensure it meets the requirements and is of high quality	Evaluate the feasibility of the approach using real-world RoPA documents.

6.3.1 Interoperability Specification for RoPA Information

This section provides the requirements specifications for DPCat based on the NeOn methodology [38]. The specification has been prepared based on the requirements gathered for ERoPA in Section 4.7.

Purpose: DPCat must provide a specification to enable the representation, collection, and transfer of RoPA information in a machine-readable and interoperable manner. It must use a consistent common model and vocabulary that facilitates the consumption and aggregation of information from multiple sources. DPCat must support the DPO's obligations to monitor GDPR compliance to identify risk and non-compliance (see Section 4.7 requirement 3.6).

Scope: The scope of DPCat is limited to modelling information relevant to the representation and use of RoPA, which can be implemented in a distributed manner in an organisation or across multiple organisations. The specification must be extensible to support the inclusion of additional information required for other GDPR compliance requirements which are not in the scope of this work (see Section 4.7 requirements 1.5, 2.1, 3.3).

Intended users: While the DPO is the primary intended user as they are the most involved with RoPA activities, the other stakeholders (see Section 6.2) based on the five use-cases identified earlier and their interactions with RoPA (see Section 4.7 requirement 2.1) are as follows: (i) U1 Data Controller, DPA regulator, certification body (ii) U2 Organisational Units (iii)U3 Data Processors (iv) U4 Data Controllers in a Joint Controllers relationship and (v) U5 DPO providing 'DPO-as-a-service.'

Intended Use: The specification is intended to support collecting and transferring GDPR accountability data, which is interpreted to mean the RoPA, which will help stakeholders complete their GDPR compliance tasks and identify risks and non-compliances (see Section 4.7 requirements 3.1, 3.5, 3.6).

Non-Functional requirements: (i) The specification should be available in English language, (ii) the specification should utilise standardised ontologies, (iii) the specification should contain cardinality rules and necessity requirements for the fields, (iv) the specification must support multiple and partial RoPA records interoperability between organisational units, departments, controllers, processors, auditors, and regulators and (v) DPCat must support conformance and provenance checking and provide the necessary utility for the user to export, analyse, and query RoPA.

Competence Questions: In chapter 4.7, the theorised requirements for an interoperability specification to enable communication in a machine-readable manner to ensure validity and conformance of RoPA with GDPR and organisational requirements are presented. The researcher, acting as a practising DPO, has identified five additional practical competencies based on

experience gathered while working with RoPA. Together, these form the competence questions for the DPCat specification to optimise data exchange validation. These are as follows:

1. Does DPCat contain the *source* of the information, e.g. department, processor? *(additional)*
2. Does DPCat *collate information* from discrete, partial, information artefacts, e.g. purpose from the department and technical measures from the processor? *(additional)*
3. Is *provenance* information present in DPCat, e.g. timestamps *(req 3.4).*
4. Does DPCat record *organisational details*, e.g. point of contact, responsible entity? *(additional)*.
5. Does DPCat maintain *distinct records*, e.g., department, processor, or temporal periods? *(req 2.1).*
6. Is DPCat sufficiently '*packaged*' for sharing RoPA record(s) with internal or external stakeholders? e.g. department to DPO or processor to the controller *(req3.1)*
7. Does DPCat support *'querying'* to retrieve partial information from RoPA, e.g., a specific period or process *(req 1.6).*
8. Is it possible to '*export*' DPCat RoPA information? i.e. to generate RoPA documentation as per requirements, e.g. GDPR Art.30. *(req3.5).*
9. Does DPCat facilitate '*customisation*'? for example, Customising information storage, retrieval, and exporting based on a variance in requirements, e.g. additional information for specific DPA templates (additional*).*
10. Can DPCat offer '*assurance*' by providing data integrity and other quality guarantees for records? *(req 3.4).*
11. Does DPCat enable '*machine-readability*' for using automation and tooling for information management? *(req 2.2).*
12. Can DPCat offer '*interoperability'* for consistency in and interpretation across stakeholders? *(req 2.1).*
13. Does DPCat provide '*openness*' for enabling adoption without lock-ins across technologies or providers? *(additional)*
14. Does DPCat offer '*extendibility*' to enable the customisation of a solution for a use-case or contextual requirements? e.g. additional terms, added information requirements **(***req 3.3).*
15. Is the DPCat RoPA record '*verifiable*'? This is required 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 *(req 3.4).*
16. Does the DPCat RoPA contain a '*sufficient granularity'* level to reflect processing activities accurately? *(req 3.2.)*

6.3.2 Knowledge Acquisition

In the knowledge acquisition stage, it is essential to understand the domain of RoPA transfer and best practices for collecting and transferring GDPR data to understand the optimal approach for collecting and transferring RoPA information.

Firstly, for the RoPA domain, each processing activity must contain all mandatory information relating to that processing activity (see Art.30) so that, in essence, each RoPA record contains all necessary compliance information, where each entry describes data processing activities within a RoPA and represents a specific context—such as a business process or data processing purpose [11].

In practice, organisations tend to prepare RoPAs based on who the controller organisation is and then populate the RoPA with all the processing activities relevant to that controller or organisation unit. This practice stems from the RoPA templates provided by regulators, where the RoPA of the document contains the controller and DPO information and conformance and provenance information, such as publisher and timestamps. Each processing activity is then represented as a singular entry ROPA Record or dataset. An Example of a well-completed RoPA provided by the DPC shows this header and line-item approach (see Figure 3). A RoPA entry may contain a header detailing the controller's name and contact details relevant to several personal data processing activities or record entries within that RoPA. This is precisely the scenario presented in the DPC example of a well-formed RoPA. In many cases, organisations maintain many RoPAs, each with its own RoPA records [11], [12]. Hence, an organisation may have multiple RoPAs by legal entity, each with its own RoPA entries. The challenge for an interoperability specification to enable the interoperability of such datasets to be generated for each processing activity and supplemented with RoPA information, such as the controller's name, whilst meeting requirements of the interoperability specification for conformance and validation and use cases (see Section 6.2) leads to a data catalogue approach to meet these requirements.

Maintaining RoPA is based upon individual RoPA records (datasets) supported with RoPA header information, which lends well to a catalogue approach. Such a catalogue approach enables the representation of diverse compliance-related data for GDPR and offers significant advantages over current methods. Some key benefits of using a data catalogue for this purpose include:

1. Data catalogues can manage diverse data types using common metadata, requiring only a small amount of data to describe processing activities.
2. Many organisations are experienced in using data catalogues, making them widely accepted in the industry.
3. User-friendly interfaces in data catalogues like CKAN [182] make them accessible to non-technical users.

4. Data catalogues support federated and distributed data processing and knowledge collection systems.
5. They have specified interoperability standards that can align with the data required for a RoPA.
6. Data catalogue models and tools can be easily extended to gather additional data for specialised datasets like a RoPA.

Using data catalogues, a data processing activity catalogue for representing diverse compliance-related data for GDPR offers significant benefits for exchanging GDPR information among stakeholders. The key benefits of Data Catalogs are that they provide a lightweight, low cost, and metadata-level integration for compliance information regarding processing activities from heterogeneous sources, thus enabling alignments between disciplinary and domain-specific metadata standards , and provides a common interoperable base for ROPA without requiring full alignment or merging all the underlying data sources [34].

6.3.3 Ontology Re-use and Engineering

The Data Catalog Vocabulary (DCAT)[150] is a W3C standard that uses RDF for publishing data catalogues on the web. For background information on DCAT (see Section 3.6). DCAT is designed to facilitate the interoperability of data catalogues, enabling datasets to be discovered and accessed across different platforms and domains [150] . DCAT helps describe datasets, data services, and catalogues in a structured, machine-readable format, making it easier for data portals, repositories, and other platforms to interoperate [34], [150].

The EU's Open Data portals utilise a data catalogue standard called DCAT-AP [154]which extends the W3C's DCAT standard for describing public sector datasets used in the open data portals. DCAT-AP enables cross-data portal search by harmonising the metadata collected and enabling common metadata collection and search for diverse national portals into a common EU portal[34], [154]. This is achieved by exchanging standard descriptions of datasets among data portals. In addition, DCAT-AP proposes mandatory, recommended, or optional classes and properties to be used for a particular application; It identifies requirements to control vocabularies for this specific application; It gathers other elements to be considered as priorities or requirements for an application such as conformance statement, agent roles or cardinalities.

DPCat has been developed by the researcher as an extension of DCAT for the representation, collection, and transfer of RoPA information in a machine-readable and interoperable manner to describe datasets so that publishers can use a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues. Compatibility with DCAT-AP is based on making DPCat usable in all DCAT-AP-based catalogue information management tools and data portals. This represents a mechanism for sharing RoPA-

related information using an EU-advocated standard and promotes the possibility of reusing existing data portal infrastructures for compliance-related purposes, such as RoPA requirements between controllers, processors, and regulators.

DPCat uses CSM RoPA to represent the information in a RoPA, which in turn uses the Data Privacy Vocabulary (DPV). DPCat represents a RoPA as a 'catalogue' of processing activities, where each activity is akin to a dataset in DCAT. Through this, a typical RoPA spreadsheet document can be represented as a catalogue consisting of multiple activities represented by each row within the spreadsheet. DPCat also enables documenting important provenance information for how, when, from where the RoPA information was obtained, and to create validation rules to ensure the RoPA information is complete and correct for use with GDPR compliance requirements. An example of this is using timestamps and publisher information to understand who is generating the RoPA information and a timestamp to identify and retrieve the most recent record, thus providing the DPO with an indicator of the up-to-date status of the processing activity.

6.4 DPCat Implementation

DPCat models the data processing activities or *entries* within a RoPA, where each entry represents a specific context—such as a business process or data processing purpose. Each processing activity is thus represented as an instance of the *RoPARecord* concept. This *RoPARecord* equates to a dataset, semantically represented by extending the DCAT(-AP) concepts. In the knowledge acquisition section, processing activity or *RoPARecord* was described as a singular processing entry and required additional data to be usable to the DPO, such as who the controller is or when it was issued. This information is presented in a RoPA which equates to Catalog in DCat-AP and uses the semantic representations of *Catalog*. This structure is presented in Table 35, with a graphical representation in Figure 23. DPCat also uses the DCat-AP concept of *Catalog* to represent a catalogue of multiple ROPA as a *ROPACatalog*. This may be required to describe a collection of RoPA documents (i.e., as a catalogue of catalogues) when an organisation has multiple RoPA documents, such as representing different temporal periods or activities or organisational units (e.g., departments).

Table 35 DPCat Definition of Terms between DCAT-AP and DPCat.

DCAT-AP Term	**DPCat Term**	**Description**
DataSet	ROPARecord	A data set 'dataset' represents an individual processing activity or 'ROPARecord.'
Catalog	ROPA	Represents a collection of entries/RoPA records.
Catalog	ROPACatalog	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).

Based on this, DPCat's modelling of RoPA is summarised as follows: 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 are associated with relevant information provided, such as the publisher (indicating who produced the record); temporal annotations, Indicating the time period represented); and annotations, such as titles and descriptions. This information enables data users to understand the provenance and source of the record, and what processing activity was active in a temporal period. A *ROPACatalog* is a catalogue for grouping together related RoPA and extends dcat:Catalog.

For everyday 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.

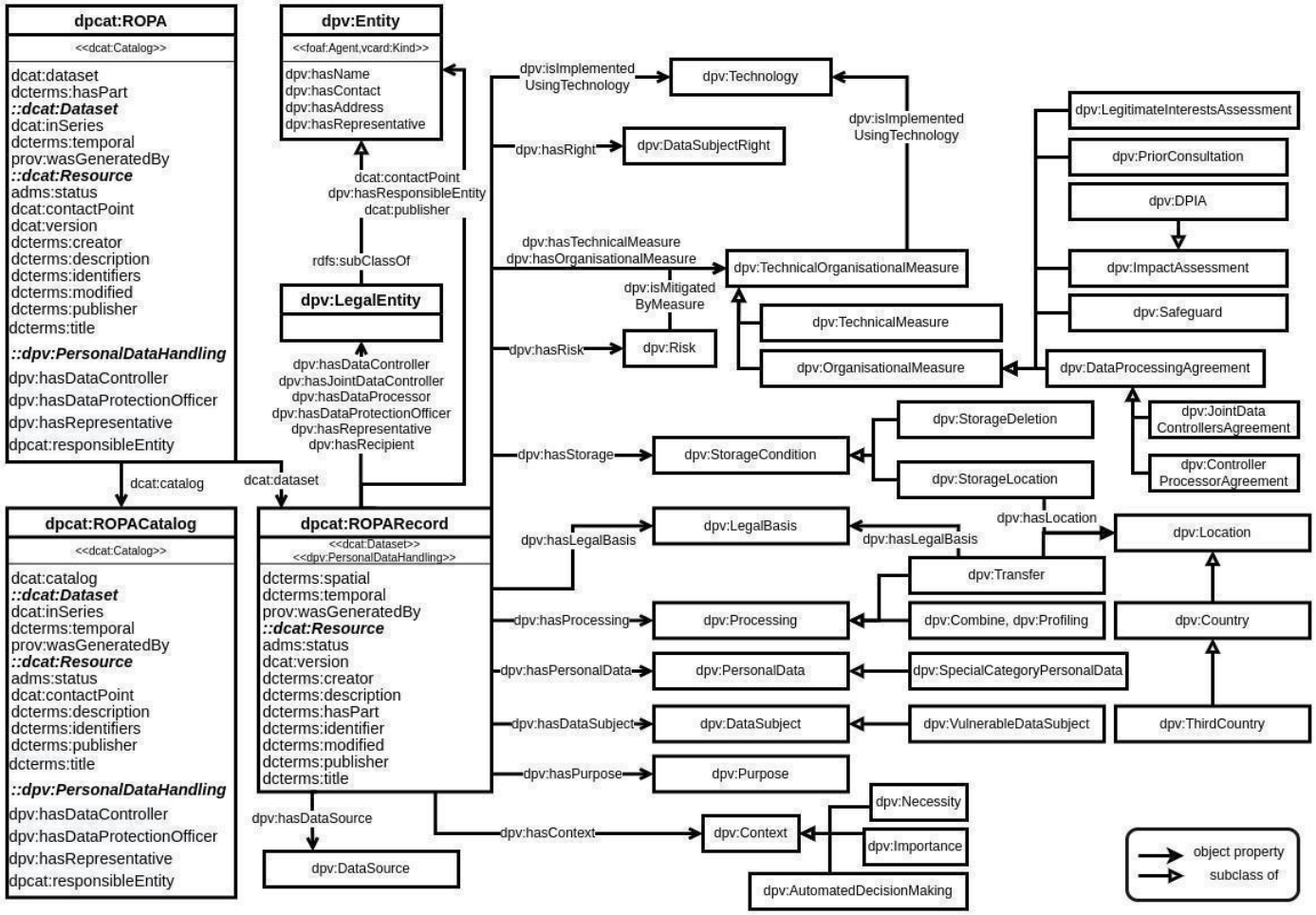

Figure 23 DPCat specification for interoperability of GDPR information.

Table 36 DPCat ROPA and ROPACatalog fields.

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

Table 36 summarises the *RoPA* and *ROPACatalog* in the DPCat specification. The 'Card.' columns refer to the cardinality of the field, and the 'Nec.' columns refer to necessity requirements for the fields, where M indicates 'Mandatory;' C indicates 'Conditional,' and MC indicates 'Mandatory Conditional,' i.e., if applicable; R indicates 'Recommended;' and O indicates 'Optional.'

ROPARecord in addition to extending dcat:DataSet, also extends *dpv:PersonalDataHandling* to associate concepts such as purposes of processing or legal bases using the relevant DPV relations as described from the CSM-RoPA mapping exercise. To ensure compatibility with DCAT and DCAT-AP requirements and recommendations, such as a publisher being a *foaf:Agent*[183], DPCat declares the relevant DPV concepts as a subclass of DCAT(-AP) specified concepts. In a *ROPARecord* instance, the concepts are coherent, i.e. all purposes apply to all personal data and are shared with all recipients. Thus, each ROPARecord represents an atomic unit of processing activity, like how each entry or row within a RoPA represents a single processing activity. A sample of the DPCat ROPARecord Specification is presented in Table 37. Refer to the online resource for the full table.[44]

[44] https://doi.org/10.5281/zenodo.14914848

Table 37 Sample of 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

The DPCat specification provides an ontology for collecting and transferring GDPR RoPA information between GDPR stakeholders while supporting the provenance and validation of interoperable datasets.

6.5 DPCat Evaluation

This section outlines the evaluation conducted to determine the effectiveness and comprehensiveness of the DPCat specification in facilitating the collection and transfer of GDPR RoPA accountability information. The evaluation includes a case study involving five common GDPR scenarios to gather and transfer GDPR RoPA compliance information among stakeholders based on the established use cases (U1-U5). The case study supports the second BIE stage of the ADR methodology.

To demonstrate the capability of DPCat to meet these competencies and use cases, a series of real-world 'typical' DPO tasks from the Data Protection Professionals survey (see Section 4.6.2) are conducted using DPCat. These tasks are referred to as Scenarios 1-5 and are as follows:

- Scenario 1: Collect RoPA Data from Organisational Unit
- Scenario 2: Validate the collected data and load RDF to GraphDB[45] triple store
- Scenario 3: Retrieve information for demonstrating GDPR Article 30 compliance

[45] https://graphdb.ontotext.com/

- Scenario 4: Generate an overview of processing for the DPO
- Scenario 5: Transfer RoPA in a graph format or as a regulator template format

6.5.1 Case Study Technical Approach

This case study validates the application of DPCat in the representation, collection, and transfer of four real-world ROPA using documents published by the European Data Protection Supervisor (EDPS). These EDPS records are first manually converted to a machine-readable format using the DPCat specification, with the data then validated using SHACL to ensure the correctness of information as per DPCat requirements, and finally, it is loaded to a GraphDB triple-store. In between these, a semantic reasoner is used to derive implicit information, such as indirect parent classes and types, based on RDFS and OWL inferencing rules to support the validation and later querying activities. At the end of these exercises, all RoPA information is collected in a single knowledge graph, which enables accessing the entire organisation's RoPA information from a single interface. A SPARQL query is then used to retrieve specific information to meet stakeholders' needs, and to export the information as a CSV (spreadsheet) or RDF (for interoperability with other stakeholders) based on specific DPA RoPA templates. This is illustrated in Figure 24. Through this case study, the practicality and feasibility of DPCat in ROPA information management is demonstrated.

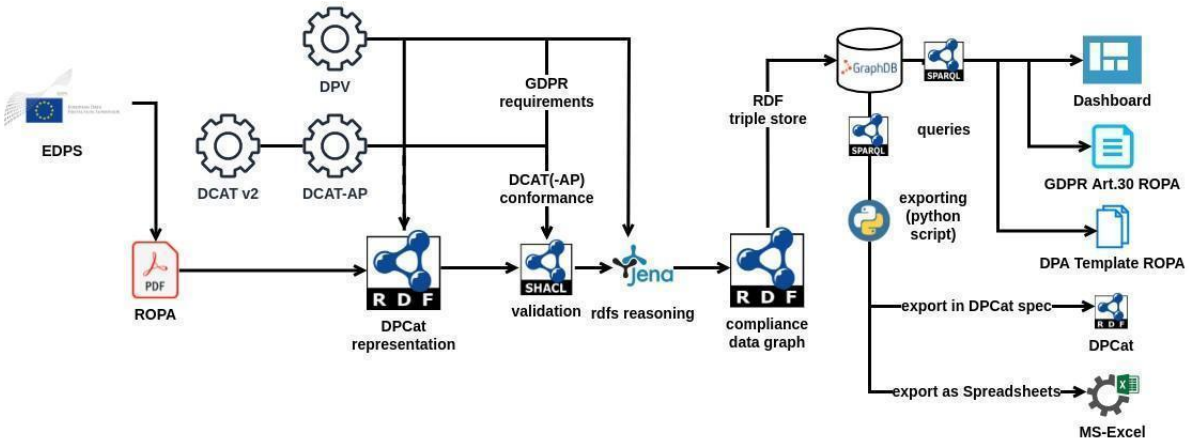

Figure 24 DPCat Case Study Technical Approach

6.5.2 Case Study Set-up

For this use-case, the real-world RoPA documents from the European Data Protection Supervisor (EDPS) were selected as the 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 publishes detailed ROPA documents based on the GDPR Art.30 requirements to meet their obligations of transparency and accountability. As of

March 2022, the EDPS has made 58 ROPA documents available in PDF format which contains information in English regarding processing operations [18]. Each document refers to a specific '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 security).

Table 38 Cross-reference of Use cases and Relevant EDPS RoPA Records.

Use Case Reference (see Section 6.2)	**EDPS RoPA Record Reference**	**EDPS RoPA Record includes this Entity**			
		Organisational unit	**Data Controller**	**Joint Controller**	**Processor**
U1: Data Controller	57		Yes		
U2: Data Controller with Internal Organisational Units	01	Yes	Yes		
U3: Data Controller with Data Processors	5	Yes	Yes	Yes	
U4: Data Controller in a Joint Controller Relationship	13	Yes	Yes	Yes	Yes
U5: DPO Overseeing Multiple Data Controllers	13	Yes	Yes	Yes	Yes

The EDPS RoPA documents were analysed to identify RoPA records covering the U1–U5 use cases detailed in Chapter 6.2 for departments, processors, and joint controllers. Four specific records (IDs 01, 05, 13, 57) were selected for the case study based on a manual selection process to ensure each use case was represented. These records are presented in Table 38. The EDPS RoPA extends beyond these four specific records, but due to the extensive labour and analysis efforts required and because the selected documents provided a sufficient breadth of processing and variability, additional documents were not integrated into the use case.

Scenario 1 Collect RoPA Data from Organisational Unit: This section describes the conversion of the EDPS RoPA records presented on the EDPS site as PDF documents[46] into an RDF format for loading to triple store (knowledge graph). Each PDF represents a single RoPA instance intended for human comprehension and lacks consistent semantics, e.g., the purpose field contains a *legal basis.* As DPCat representations are machine-readable information, they require explicit separation of concerns to represent the processing activities correctly. Therefore, where the EDPS RoPA contains combinations of different processing activities within the same description, the researcher should manually represent them as separate processing activities in DPCat. For example, EDPS RoPA Record 13 (see Appendix H) specified two processors, are interpreted as separate *ROPARecord*

[46] ttps://www.edps.europa.eu/about/data-protection-within-edps/records-register_en

instances for each processor to indicate the separation of concern in the controller's communication and data governance. The whole collection of documents and RDF graphs were then expressed as part of a single *ROPACatalog* instance reflecting the published set of records on the EDPS website. The manually created RDF graphs were then enhanced using the Apache Jena [184] RDFS reasoner to create a 'complete graph' to simplify querying and validation. The RDFS reasoning is sufficient to obtain the expansion of subclasses and sub-properties within the graph rather than generate inferences using an OWL reasoner. For storing the information and offering a querying interface, the GraphDB triple store was used as it is a freely available triple-store compliant with relevant standards (e.g. SPARQL) and has several features for convenience, e.g. friendly interface, integrated reasoners, SHACL validation.

A sample RoPA record based on the EDPS ID 05 document and defined using DPCat is presented in Listing 1. This record provides information regarding the selection and management of interim staff. Some examples of the information contained in the record are that it was published by the HRBA on 03/11/2021. The record provides a contact point of the HRBA and uses the legal basis of contract.

Listing 1 Representation of ROPA and ROPARecord for EDPS Document.

```
@prefix edps: <https://w3id.org/dpcat/examples/EDPS/vocab#> .
@prefix : <https://w3id.org/dpcat/examples/EDPS/05#> .
: a dpcat:ROPA ;
   dct:title 'Selection of staff'@en ;
   dct:description 'The purpose of the processing ...'@en ;
   dct:created '2021-11-03'^^xsd:date ;
   dct:identifier '05'^^xsd:string ;
   dct:publisher edps:HRBA ;
   dcat:contactPoint edps:HRBA ;
   dcat:dataset :05-1, :05-2, :05-3, :05-4 .
:05-1 a dpcat:ROPARecord ;
   dct:title 'Selection and management of interim staff'@en ;
   dct:description 'Selection and management of interim staff'@en ;
   dct:created '2022-02-16'^^xsd:date ;
   dct:publisher edps:HRBA ;
   dcat:contactPoint edps:HRBA ;
   dpv:hasDataController edps:EDPS ;
   dpv:hasResponsibleEntity edps:HRBA ;
   dpv:hasPersonalData edps:InterimAgentLastName,
      edps:InterimAgentFirstName, edps:InterimAgentCV ;
   dpv:hasPurpose edps:MonitoringOf7YearRule, edps:PreparationOfEmploymentContracts,
edps:ExecutionOfContract ;
   dpv:isImplementedUsingTechnology edps:SYSPER ;
```

```
dpv:hasRecipient edps:Managers, edps:HRBA ;
dpv:hasLegalBasis dpv:StaffContract ;
dpv:hasTechnicalOrganisationalMeasure edps:AuthorisedPersonnelNeedToKnowBasis ;
dpv:hasDataSource edps:RandstadBelgium, edps:Daoust ;
dpv:hasStorage [
  a dpv:StorageCondition ;
  skos:editorialNote 'The data ... managed by Sysper'@en ;
  dpv:hasDuration [
    a dpv:StorageDuration, time:Duration ;
    dpv:hasPurpose edps:EnsureReconstructionOfHistoryBySYSPER ;
    dpv:hasPersonalData edps:DataRelatingToSYSPERService ;
    time:numericDuration '5'^^xsd:decimal ;
    time:unitType :unitYear ; ] ] ...
```

Scenario 2: To Conduct a Validation Check on a RoPA Record: This section describes the validation of the transformed EDPS RoPA file. The ability to validate RoPA information is a critical component of ERoPA. This process involves verifying and validating the generated RDF graphs using Shapes Constraint Language (SHACL) constraints provided with DPCAT-AP specifications to ensure data correctness according to DCAT and DPCAT-AP specifications, e.g., publishers being of type foaf:Agent. In this use case, the open-source and freely available TopBraid [156] SHACL tool is utilised for executing the constraints.

Here, it is important to assert that such validation shapes are not readily available, for example, for use with DPV concepts. Consequently, there may be more than one way to define a 'shape' for a given scenario, often at arbitrary levels of complexity, which prevents 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 location can be modelled in terms of *dpv:hasLocation* of *dpv:DataTransfer*. However, the Data Transfer instances could be used at any arbitrary node within the graph, making it challenging to define follow-up constraints such as the recipient of that transfer and its location. This indicates a need for further constraining the use of ontologies used here – namely the DPV – to have a known set of representations for expressing specific scenarios through which SHACL shapes can be provided to optimise ERoPA validation of RoPA information. This may be achieved by identifying use cases for each concept's use and defining specific SHACL shapes for how that information should be expressed using DPV, and ideally providing this information alongside DPV's concepts to ensure their use is consistent across use-cases. For the purposes of this use-case, SHACL shapes were created based on an interpretation of how the concepts should be used.

Scenario 3 To output an Article 30 format RoPA for Regulator: In this scenario a 'typical' DPO task is conducted to demonstrate the utility of DPCat. The task is to generate a RoPA as a set of information required by GDPR Article 30. Listing two shows the query used for this task, while Table 39 demonstrates the query's output.

Listing 2 SPARQL Query to Generate Article 30 RoPA.

```
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 } }
```

Table 39 Query Output of Article 30 Format RoPA for Regulator.

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 Speedwell operators		
Financial Transactions	Budgetary commitments	Staff members	Bank Account	Local Profile Manager		
Financial Transactions	Payments	Staff members	Bank Account	The EDPS Financial team		

Scenario 4: To generate an overview of processing for the DPO: In this scenario, the task is to retrieve information required to generate an overview of processing for the DPO. This task demonstrates the potential for DPCat to help create internal reports or dashboards based on ROPA information. The query for this task is displayed in Listing 3, where relevant information concerning the organisation's processing activities and relationships with external entities is retrieved from *ROPARecord* instances to enable the DPO to understand the organisation's activities and external engagements at a high abstract level. The output of the query is shown in Table 40. Due to the nature of DPCat as a semantic model and the use of a triple store, it is a trivial task to further explore

each activity e.g. by using the SPARQL DESCRIBE query or by creating detailed views in a dashboard type interface for the DPO,

Listing 3 SPARQL Query for an Overview of Processing for DPO using DPCat.

```
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 } }
```

Table 40 Query Results for an Overview of Processing for DPO 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

Scenario 5 Transfer - graph and spreadsheet regulator template: DPCat provides an approach for information exchange and data governance within and between GDPR stakeholders. As a first example, this information is exported as DPCat-defined catalogues using SPARQL CONSTRUCT queries to export information as an RDF graph or through other formats supported by the triple-store such as CSV spreadsheets, to share with entities, provide inputs to tools, or to be stored as

backups or logs. The exported data can also facilitate the exchange of RoPA information as RDF graphs between entities if both support the DPCat specification, and if not – then still as CSV spreadsheets containing human-readable documentation which is widely supported. The SPARQL query and the resulting graph can be viewed online[47].

Another example of such information interoperability is the automation of a DPO that manually manages information in a spreadsheet, for example, through a Python script to execute SPARQL queries and export 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, additional scripting using Python is helpful to replicate the DPA template's structure and contents and operate over the more complex XLSX format that supports tabs within spreadsheets.

6.5.3 Case Study Results

This section discusses the results of the DPCat evaluation. The results are presented in Table 41. The findings are prepared based on observations gathered throughout the case study.

Table 41 Summary of Outcomes from DPCat Case Study Scenarios.

Activity	**Outcome**
Scenario 1: collect RoPA data from the organisational unit.	Manual conversion of PDF documents completed Organisational data successfully represented in RDF format Controlled vocabulary created RoPA data successfully transferred from the organisational unit to the knowledge Graph
Scenario 2: validate the records and load RDF to GraphDB triple store.	SHACL shapes created for validation of RoPA record Successfully identified non-conforming RoPA records There is a need for predefined SHACL shapes to be defined to which datasets must conform.
Scenario 3: output article 30 format RoPA for Regulator.	Developed SPARQL query to generate Article 30 format RoPA Successfully generate Article 30 RoPA
Scenario 4: generate an overview of processing for the DPO.	Successfully provided a query that displayed relevant information concerning the organisation's processing activities and relationships with external entities retrieved from *ROPARecord* instances
Scenario 5: Transfer - graph and spreadsheet regulator template.	Successfully exported DPCat-defined catalogues using SPARQL CONSTRUCT queries to retrieve related information as an RDF graph for exchange Successfully executed SPARQL queries and exported results into an MS Excel (.xlsx) document based on DPA ROPA templates

[47] https://github.com/Paul-Ryan76/DPCat

6.5.4 Analysis

This section concerns the capacity of DPCat to meet the competency questions required of the interoperability specification (see Section 6.3.1). The case study shows that the RoPA records represented using the DPCat specification are machine-readable and aligned with an open, standardised approach to interoperability (see Section 4.7 requirements 2.1 and 2.2). DPCat has been successfully employed to semantically represent four records from the organisational units taken from the EDPS RoPA. These records were originally in PDF format and designed solely for human consumption, with no preconception of being required to be machine-readable documents. They were successfully converted to RDF using CSM-RoPA and expressed as a machine-readable RoPA using DPCat, which was then used to simulate the tasks of the DPO such as generating information required for GDPR article 30.

Applying DPCat to real-world ROPAs has revealed inherent challenges in creating semantic representations. The input data is either incomplete or loosely structured, which does not align well with the strict structure required by machine-based tools. In the case of the EDPS RoPA for the four modelled records, this required the construction of a controlled vocabulary. This task could become a significant challenge for a large unstructured RoPA where the organisation needs better data governance. This issue could be addressed if the organisation develops a separate registry of controlled vocabulary and improves data governance capabilities. However, organisations approach significant debates that lack structured data collection methods. Organisations' different approaches to modelling create obstacles to using DPCat as a standard information representation mechanism.

The case study demonstrates that the DPCat specification supports conformance and provenance checking when collecting GDPR RoPA data. Using SHACL shapes to validate RoPA records provides a valuable tool for DPOs to identify nonconforming data sets. The case study used a SHACL validation where publishers must be of type foaf:Agent. This test was successful. The case study has indicated a need for predefined SHACL shapes to be defined to which datasets must conform.

The case study demonstrates the utility of DPCat for exporting and querying RoPA information. Scenarios 3-5 use queries to export data in many forms, such as standard article 30 RoPA, an RDF graph file, or a spreadsheet. This enables the successful querying and transfer of GDPR compliance data with stakeholders. The EDPS data set used for the case study provides examples of meeting use cases U1-U5. Each of these use cases was successfully modelled using the test scenarios. The ability to transfer data in Article 30 form, regular template, RDF graph form, or

a spreadsheet (see scenarios 3-5) would enable the external DPO to review RoPA data efficiently (use case U5).

The DPCat satisfies fourteen competency questions from the requirements specification, partially satisfies one, and does not satisfy one question, as shown in Table 42. The 'Status 'column uses a key where C = completely met, P = partially met, and N not met. The Column 'Scenario' relates to the five scenarios (see Section 6.5.2).

Table 42 Extent that DPCat meets the Competence Questions.

Question no.	**Competence Question**	**Status**	**Scenario**	**Outcome of Evaluation**
1	Does DPCat contain the *source* of the information?	C	All	The publisher, responsible entity, processor, and controller are identified.
2	Does DPCat *collate information* from discrete, partial, information artefacts, e.g. purpose from the department and technical measures from the processor? *(additional)*	C	All	The four EDPS RoPA records were successfully loaded
3	Is *provenance* information present in DPCat, e.g. timestamps? **(req 3.4)**	C	All	Each RoPA record contains times stamps, the publisher, the responsible entity, the processor, and the controller.
4	Does DPCat record *organisational details*, e.g., the point of contact and the responsible entity? *(additional)*.	C	All	The point of contact and the responsible entity are identified
5	Does DPCat maintain *distinct records*, e.g., department, processor, or temporal periods? **(req 2.1)**	C	All	Each loaded record is uniquely identifiable. Each RoPA record has its created data and all required identification information.
6	Is DPCat sufficiently '*packaged*' for sharing ROPA record(s) with internal or external stakeholders, e.g. department to DPO or processor to the controller? **(req3.1)**	C	All	Each RoPA record contains all necessary data for sharing
7	Does DPCat support *'querying'* to retrieve partial information from ROPA, e.g., a specific period or process? **(req 1.6)**	P	3,4,5	Scenarios 3-5 uses a query to gather and export a RoPA. Recommend further testing for a partial record.
8	Is it possible to '*export*' DPCat RoPA information, i.e. to generate ROPA documentation as per requirements, e.g. GDPR Art.30? **(req3.5)**	C	3,	Scenario three fully meets the requirement, in which a RoPA is exported as an Article 30 form or a RoPA regulator template.
9	Does DPCat facilitate '*customisation*' for example, Customising information storage, retrieval, and exporting based on a variance in requirements, e.g. additional	N	Not yet met	DPCat uses GDPR concepts from CSM-RoPA and interoperability concepts from DCat. Whilst catalogues can accommodate a need for additional information, this has not been identified in this case

Question no.	Competence Question	Status	Scenario	Outcome of Evaluation
	information for specific DPA templates? (additional*)*			study. This requirement remains unmet.
10	Can DPCat offer '*assurance*' by providing data integrity and other quality guarantees for records? **(req 3.4)**	C	2	In Scenario 2, the RoPA records are successfully validated to meet specific quality requirements.
11	Does DPCat enable '*machine-readability*' for using automation and tooling for information management? **(req 2.2)**	C	All	Four EDPS PDF documents are converted to RDF and loaded to the knowledge graph. These records are checked for quality and integrity to meet the information management requirements (such as publisher being mandatory)
12	Can DPCat offer '*interoperability'* for consistency in and interpretation across stakeholders? (**Req 2.1)**	C	All	The four RoPA records were successfully loaded without any interpretation challenges (example scenario 3 RoPA export)
13	Does DPCat provide '*openness*' for enabling adoption without lock-ins across technologies or providers? *(additional)*	C	All	The code and tools used for this case study are free without lock-in.
14	Does DPCat offer '*Extendability*' to enable customisation of a solution for a use-case or contextual requirements, e.g. additional terms, new information requirements? **(req 3.3)**	C	All	DPCat uses the GDPR concepts from CSM-RoPA supplemented with concepts from the DCAT ontology. These terms have proved sufficient for this case study for all scenarios.
15	Is the DPCat RoPA record '*verifiable*'? This is required to support information management through validating information in terms of correctness and completeness. **(req 3.4)**	C	2	In Scenario 2, the RoPA records are successfully validated to meet specific quality requirements.
16	Does the DPCat RoPA contain a '*sufficient granularity'* level to reflect processing activities accurately? **(req 3.2)**	C	All	All scenarios present RoPA records containing sufficient granularity to meet legal obligations.

Based on the analysis of the competence questions, fourteen questions are completely met, and two questions are partially met or not met. Competence number 7 is 'Querying,' which involves retrieving partial information from ROPA, such as a specific period or process. Although Scenario 3-5 uses a query to gather and export a RoPA, further testing is recommended for retrieving a partial record, particularly for selected SPARQL queries that a DPO may require.

The second competence, number 9, is 'Customization,' which requires customising information storage, retrieval, and exporting based on variations in requirements, such as additional information for specific DPA templates. This competence is assessed as not yet met because DPCat uses GDPR concepts from CSM-RoPA and interoperability concepts from DCat, meeting all the needs of the EDPS RoPA. While catalogues can accommodate the need for

additional information, this has not been identified in this case study. Therefore, this competence remains conceptually met.

6.6 Learnings from the Development and Evaluation of DPCat

This section takes the findings of the case study evaluation, together with feedback gathered from technologists and legal experts at conferences and presentations where DPCat was presented and reviews of publications on DPCat (see Section 6.7).

Many positives were identified in the DPCat use cases, with the scenarios and use cases being met successfully. DPCat successfully met fourteen of the sixteen competence questions for an interoperable RoPA. The capability of DPCat to support the conformance and validation of 'Provenance,' 'Assuring,' and 'Verifiable' provides a key benefit to DPOs in ensuring that the RoPA contains the necessary integrity to support GDPR compliance.

From the case study, the skills and technology required for ERoPA became evident. The manual, unstructured nature of documents (such as the EDPS PDF documents) creates challenges for machine-readable documents. This required manual input to assign the relevant metadata to the documents and took a sizeable manual effort and a degree of technical and GDPR legal skills to identify concepts within the document. This is typified whereby an important field like legal basis is found in the purpose of processing column. This would not have been anticipated or expected.

While the DPCat specification was developed and used to express RoPA in RDF, the need for additional tools to support the case study became evident. Tools such as SPARQL and SHACL were required to meet the competence questions of DPCat and meet the requirements for an ERoPA. The technology needs to support ERoPA goes beyond CSM-RoPA to represent a RoPA and DPCat to enable the collection and transfer of RoPA. In practical terms EROPA requires SHACL shapes for conformance checking and SPARQL queries is required for exporting and analysing the RoPA. In terms of these SHACL shapes, these need to be defined

The information requires consistency or foreknowledge regarding how the data is structured or 'shaped.' Without this, the resulting SPARQL queries and SHACL shapes can be complicated to express or become complex. Enforcing the consistency of the underlying information is vital to ensure the consistency of DPCat implementation, especially for information exchange. 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.

6.7 Conclusions

This chapter outlines the development and evaluation of a GDPR RoPA Interoperability specification. The goal is to facilitate the representation, collection, and transfer of GDPR accountability information among data protection stakeholders. This chapter specifically focuses on the second BIE stage of ADR methodology. It aims to address two research sub-questions: (1) RSQ2.c - developing a specification for the interoperable exchange of RoPA information based on semantic-web standards and best practices, and (2) evaluating the extent to which machine-readable RoPA supports demonstrating GDPR accountability.

The NeOn Methodology was employed to design and implement the DPCat interoperable specification. The specification is then evaluated using a use case where DPCat supports standard DPO tasks. The evaluation confirms that DPCat can successfully support the representation, collection, and transfer of RoPA and provide practical support for demonstrating GDPR accountability. The chapter also provides evidence for the practicality and feasibility of supporting the DPO in demonstrating GDPR accountability while identifying limitations of DPCat and areas for future work. The evaluation confirmed that DPCat meets the requirements for an interoperability specification to support the exchange of ERoPA. The successful completion of the case study scenarios demonstrates that DPCat can support demonstrating GDPR compliance.

6.7.1 Peer-Reviewed Publications

The development of the DPCat specification has been published in two journals (see Table 43 below) and at the SEMANTiCS 2021 conference in Amsterdam in September 2021 for professionals and academics from various domains, including a special track on semantic regulatory technologies.

Table 43 Publications and conference presentations concerning DPCat

Publication	**Venue**	**Citations (Feb 2025)**
Support for enhanced GDPR accountability with the standard semantic model for RoPA (CSM-RoPA)	Springer Nature Computer Science Journal Paper	11
Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-Related Information for GDPR Using DCAT-AP and DPV	SEMANTiCS 2021, the 17th International Conference on Semantic Systems, Amsterdam, Sept 2021	5
DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR	Journal of Information, *Information* 2022,	11

The work was positively accepted, and valuable feedback was gathered to develop the DPCat specification further. DPCat was also presented at ‘ADVANCE 2023: Generative AI, Large Language Models and Beyond: Strategies for Adoption and Integration’ on Wednesday, June 7th, 2023, at Trinity College Dublin.

7 The ERoPA Approach and Upsilon Case Study

7.1 Chapter Overview

This chapter provides an overview of the full ERoPA Approach and a validating case study. The ERoPA deployment case study contains the CSM-RoPA ontology for RoPA representation (see Chapter 5), enhanced with the DPCat RoPA interoperability specification (see Chapter 6) and all ERoPA Approach component tools and methods, such as RDF conversion, a triple store, quality assurance control, and query tools, in the third BIE cycle.

The 'Upsilon'[48] ERoPA Deployment Case Study evaluates the deployment feasibility and capability of the ERoPA Approach to deliver accountability compliance in real-world organisations by (i) deploying the ERoPA Approach into a sandbox environment mimicking Upsilon's technical and organisational structures, (ii) gathering and analysing observations from the Upsilon ERoPA sandbox deployment process, (iii) gathering the views of experts as potential end users, (iv) analysing the extent to which the Upsilon ERoPA deployment supports the requirements of a regulator-supplied accountability tracker and (v) provides guidelines for organisations deploying ERoPA. This chapter evaluates the extent to which the ERoPA Approach enables the **ERoPA system capabilities** (see Section 7.2.3) in an organisation.

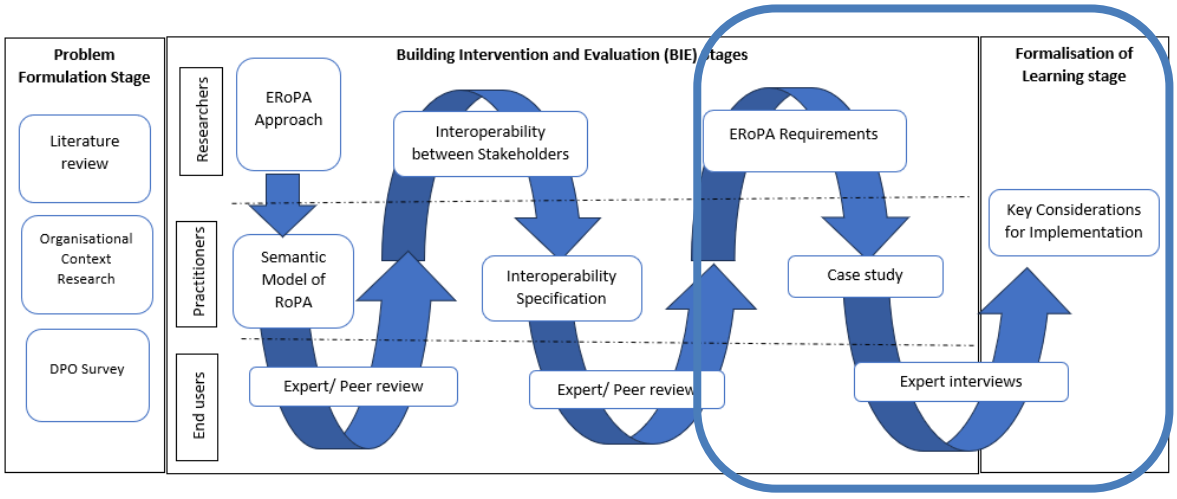

Figure 25 Action Design Research showing the Stages covered by this chapter.

This chapter addresses RSQ3 to establish the extent to which the application of the ERoPA supports implementing GDPR accountability. The chapter also addresses RSQ4 to identify the key considerations for organisations implementing the ERoPA Approach by providing a structured

[48] Upsilon is the fictional name given in this thesis to a real organisation that was studied as part of this work. It includes over 1000 employees and has a group structure including multiple DPOs in its operating companies

approach to support organisations in implementing ERoPA. The chapter addresses the third ADR Building, Intervention, and Evaluation stage and formalises the learning stage of ADR, where the findings from the case study are combined with all the findings gathered through this thesis to provide guidelines for deployment (see Figure 25).

Description of Chapter: Section 7.2 presents the ERoPA Approach and provides an overview of its key components of ERoPA. 7.3 introduces the Upsilon ERoPA deployment case study plan, 7.4 the case study design, 7.5 the data collected in the case study, and 7.6, an analysis of the data collected. In section 7.7, the findings of the ERoPA case study are presented. Section 7.8 distils the case study findings into deployment guidelines for organisations deploying ERoPA. Section 7.9 describes the conclusions gathered from this chapter.

7.1.1 ADR Roles

The case study deployment utilises designated ADR Roles for the BIE 3 stage (see Table 44). The first of these roles, the researcher, a practising DPO with experience in conducting the planning and implementation of the ERoPA, conceptualises the case study deployment. The second role, the Practitioner role, is met by the Upsilon Data Protection team, who carry out all the deployment tasks and gather deployment observations. The ADR role is completed by Data protection experts who provide feedback on the ERoPA Upsilon deployment.

Table 44 ADR Role Assignment for BIE 3

Role	Assignment
Researcher	The Thesis Researcher (a practising DPO)
Practitioners	Upsilon Data Protection team
User	Expert Interviews

The formalisation of learning stage of ADR is completed by the researcher who gathered the findings from the State of the Art review (see Section 2.6), Survey of Data Protection Professionals (see Section 4.6.2), CSM-RoPA implementation (see Section 5.5), DPCat-use cases (see Section 6.6) and the ERoPA deployment Case study (See Section 7.7) to provide a framework for implementing the ERoPA Approach (see Section 7.8.4)

7.2 The ERoPA Approach

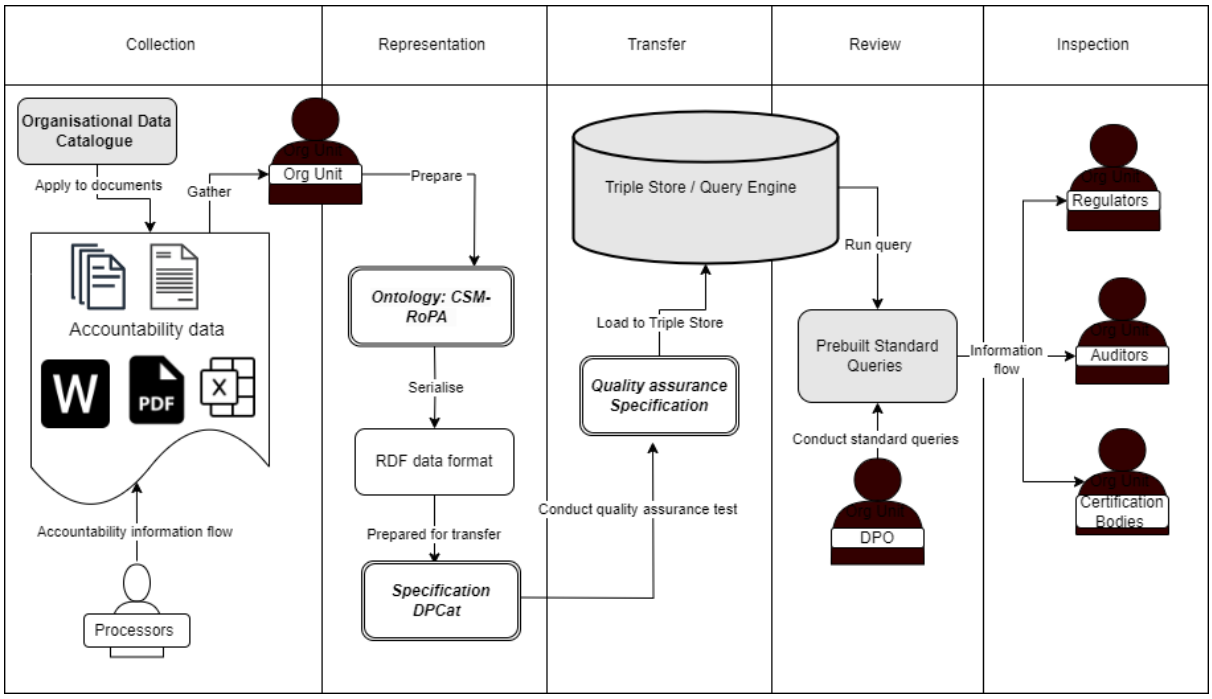

Figure 26 Overview of the Tools and Methods Used in the ERoPA Approach.

The ERoPA Approach involves five key stakeholder activities, which are as follows: (i) the **collection** of RoPA information from stakeholders such as organisational units or data processors, which involves the gathering of RoPA information from all sources (such as Data Processing Agreements and Privacy Notices and , Policies and Procedures for example) (ii) the **representation** of this information by way of conversion to a machine readable format (iii) the **transfer** of the information between stakeholders whilst complying with provenance and conformance rules (iv) the **review** of this information by the DPO using tools to support typical DPO and (v) the **inspection** of this information by regulators, auditors and certification bodies, by providing RoPA data in a human readable or machine readable format.

To achieve this GDPR compliance information flow, the ERoPA Approach involves seven key methods and tools, referred to as 'components.' These components are divided into two categories: (i) tools and methods developed explicitly for the ERoPA Approach and (ii) other tools and methods used to deploy the ERoPA Approach. The stakeholder activities tools and methods used to support stakeholder information flows in the ERoPA Approach are presented in Figure 26 and are described below.

7.2.1 Tools and Methods:

The tools and methods of ERoPA that have been developed in this research are as follows:

- The CSM-RoPA ontology was developed to represent all the GDPR concepts for RoPA (see Section 5.1).
- The DPCat specification was developed to support collecting and transferring GDPR RoPA information between GDPR stakeholders (see Section 6.1).

  The following tools and methods have been used to deploy the ERoPA Approach:
- A Data Catalogue is used to organise, store and manage datasets. Data catalogues support metadata management for describing datasets effectively. They offer robust search facilities and provide APIs for easy integration. Data catalogues are extensible and facilitate the publishing of data sets. (For more information on data catalogues refer to Chapter 2.2). The catalogue in the DPCat specification is expressed as the DCAT information system records each processing activity as a dataset, and a RoPA or header information represents a predefined structure for exchanging information. When data is presented in the DPCat catalogue structure, it benefits that a data catalogue provides such as the ability to publish and query datasets using specialised tools in a triple store.

  or managing and publishing datasets
- A tool for Data Conversion to RDF was required to create a schema to support the conversion of data presented as a spreadsheet into RDF format. This was achieved using the OntoRefine tool to create a schema and apply the schema to a spreadsheet to convert the data into RDF format suitable for loading to the knowledge graph.
- A Triple Store was required to store RDF triple data. The Graph DB tool provides inbuilt tools to assist in importing RDF into the knowledge graph triple store. These imports can come as a URL link, an upload file or a clipboard text string directly entered the upload tool. This gives the user an easy-to-use interface for loading RDF data to the knowledge graph. Graph DB also provides SPARQL query as an inbuilt feature.

  Graph DB can also be explored using the GUI to visualise the knowledge graph or specific parts.
- A Data Quality Assurance Specification was required for conformance and provenance checking to ensure that input to RoPA provided by stakeholders met data quality rules. The SHACL data shapes ontology was used to create preferred graph shapes, which were used to identify data quality issues and non-compliances.
- A Query Tool was required to support the DPO in conducting typical data protection tasks. The query engine SPARQL was used for these processes. SPARQL forms part of the triple store and is used to process queries. A set of standard queries based on typical DPO tasks

were created. These queries can be used for tasks such as the export of the organisational RoPA to a spreadsheet form or to extract a particular processing period that is valid at a specific period. Together, these components make up the ERoPA Approach to GDPR compliance.

7.2.2 The Role of Stakeholders in the ERoPA Approach

This section describes the six distinct stakeholder groups and their roles in GDPR compliance. The key stakeholders are processors, controllers, auditors, certification bodies, regulators, organisational units, and data protection officers. The key activities they conduct are as follows:

- Collection: gathering and preparing GDPR RoPA compliance information.
- Representation: presenting this GDPR compliance information in a machine-readable format.
- Transfer: the exchange of GDPR information between stakeholders in a machine-readable format
- Review: the activities concerned with monitoring GDPR compliance by the DPO.
- Inspection: the activities conducted by external parties (outside the organisation) concerned with the review of GDPR compliance.

Table 45 shows each stakeholder's role with ERoPA, Section 3.3.3 provides more detail on each stakeholder's role in RoPA, while Section 7.2 describes the many use cases of data transfers between stakeholders.

Table 45 Role Played in ERoPA by Stakeholders.

GDPR stakeholders	Collection	Representation	Transfer	Review	Inspection
Organisation Units	✓				
Processors	✓	✓	✓		
Controllers	✓	✓	✓	✓	
Data Protection Officer	✓	✓	✓	✓	
Auditors and certification bodies				✓	✓
Regulators					✓

7.2.3 System Capabilities

This section describes the capabilities organisations can do because of deploying the ERoPA Approach. A capability is an artefact to perform a coordinated task, utilising organisational resources, to achieve a particular result [3]. The organisational capabilities are gathered from two primary sources. Firstly, building on the work of Labadie and Legner [41], a capability model for data management in GDPR is prepared and extended with 'Maintain ROPA' system capabilities using the ERoPA Approach—highlighted with bold text and red borders [18] [41].

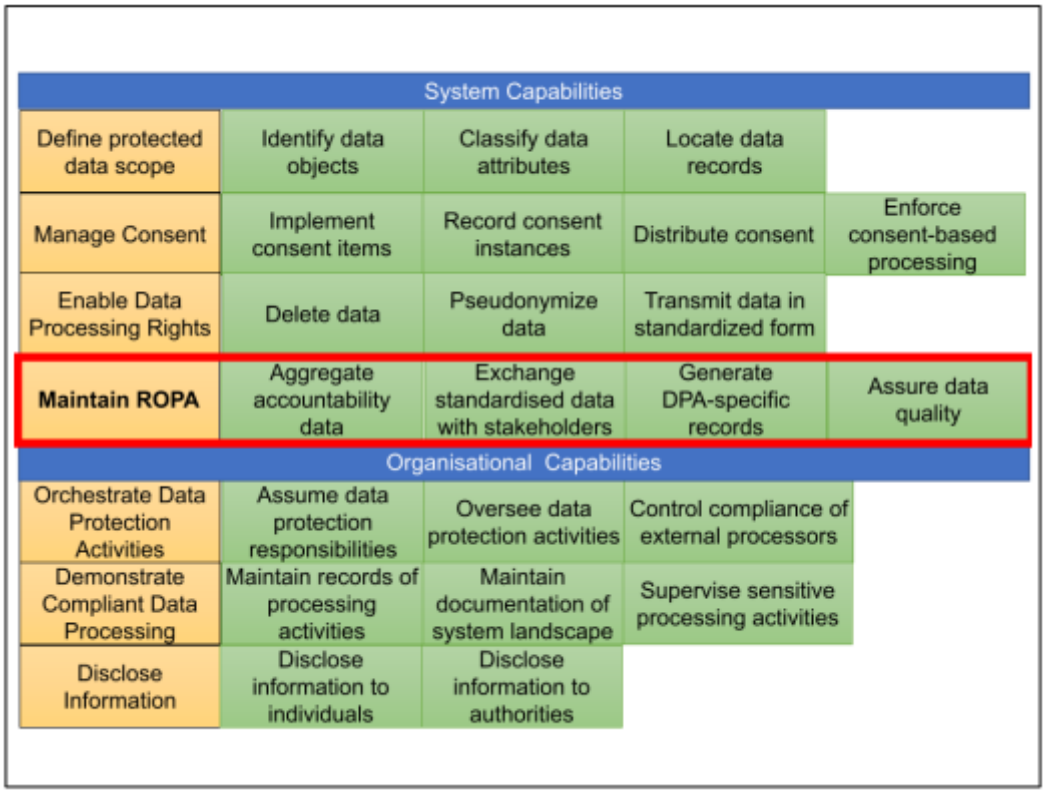

Figure 27 Extended Capability Model for Data Management in GDPR [18] .

The capabilities from Figure 27 are extended with the ERoPA requirements from Chapter 4.7 to extend the capabilities for ERoPA (see red box for extended capabilities). These extended capabilities for ERoPA are presented in Table 46.

Table 46 ERoPA System capabilities and specific requirements for ERoPA

ERoPA System Capabilities	**Capability Model for Data Management** [18]	**Capability gained from ERoPA (based on ERoPA requirements, Section 4.7)**
Provides for the Comprehensive Representation of Personal Data Processing Activities	Aggregate accountability data	Requirements 1.1, 1.2, 1.4, 1.5, 3.2, 3.3
Enables Standardised RoPA Information Transfer	Exchange RoPA data with stakeholders to enable standardised RoPA information transfer	Requirements 1.3, 2.1, 2.2, 3.5
Supports Compliance checks & validation	Assure data quality	Requirements 3.4, 3.6
Provides GDPR Reporting and Review Tools for DPO	Generate records for regulators	Requirements 1.6, 3.1

The ERoPA capabilities are presented in the case study using the following tools and methods:

1. Use CSM-RoPA/DPCat to gather and aggregate accountability data from heterogeneous sources and organisations to represent personal data processing activities comprehensively.
2. Use DPCat to define IT infrastructure (data catalogues) and standard data formats (DPCat) to exchange RoPA data **with stakeholders to enable standardised RoPA information transfer**.
3. Use a triple store to store the GDPR information from which typical DPO queries are run to **generate records for review** and provide GDPR reporting tools for DPO.
4. Use SHACL rules and the DPCat specification to **assure data quality** consistency and integrity to support compliance checks and validation.

Chapter 5 used the ERoPA Approach to represent a standard Article 30 Record of Processing Activities (RoPA) using test data with the CSM-RoPA ontology. Chapter 6 introduced and supplemented the ontology with the DPCat interoperability specification to represent four records sourced from the European Data Protection Supervisor (EDPS). This chapter demonstrated the representation of these four RoPA data records and the validation and loading processes to the GraphDB triple store. It also covered retrieving this information, demonstrating how it ensures compliance with GDPR Article 30. Also, the chapter presented the generation of RoPA outputs for the Data Protection Officer (DPO) and Transfer RoPA, formatted in both graph presentations and regulator template formats.

The case study in this chapter builds upon the previous chapters to gather and represent data in a complex real-world situation. It extends the work of Chapters 5 and 6 by (i) scaling up the processing from one entity with four RoPA records to a multi-entity organisation with over 3 distinct companies, 14 organisational units, 21 controllers, 75 processors and 246 processing records, which represent various organisational units, joint controllers, data controllers, and processors; (ii) incorporating organisational data catalogues and schemas to gather the RoPA information effectively; and (iii) simulating real business activities with updates to processing records.

This case study encompasses all the ERoPA capabilities, as the case study requires comprehensive gathering, aggregation, and representation of real-world data. The data must be generated and transferred by organisational stakeholders in a standardised format, ensuring consistent quality. Subsequently, this data must be used by the Data Protection Officer (DPO) to conduct the typical responsibilities associated with their role. The next section describes the case study deployment of ERoPA in the Upsilon organisation.

7.3 Upsilon Deployment Case Study

This section provides an overview of the case study plan, including the case study research question, the purpose, the unit of analysis, and how the data will be collected and analysed. The case study utilises Yin's Case Study theoretical framework [155] to guide this case study.

Case Study Research question: What are the key considerations for organisations deploying ERoPA?

Case Study Purpose: In this case study, an EROPA deployment is conducted in a real work organisation at TRL 5 (Technology validated in a relevant environment). The main goal is to gather findings from the ERoPA deployment through deployment observations, expert feedback, and accountability verification observations to verify the ERoPA Approach and develop deployment guidelines.

The unit of analysis consists of the business units and partners of the Upsilon company involved in ROPA-related data protection activities. The phenomenon being studied in depth is the deployment of the ERoPA Approach in a real-world context, being the Upsilon organisation. The analysis of the deployment of the ERoPA Approach on data-related activities carried out by business units and partners of the Upsilon company will be the focal point of the case study.

The Upsilon organisation is a large organisation that employs more than one thousand personnel. It has three distinct companies with fourteen organisational units. Many of these organisational units are centralised services such as Human Resources and Information Technology, which provide services to other companies within Upsilon. There are seventy-five data processors

providing data processing services to Upsilon, and there are twenty-one data controllers where Upsilon function as a data processor on their behalf. Upsilon operates 58 IT systems (see Table 48 for more information).

Data Collection Procedures. This research involves three data collections as described above. For the deployment observations, the researcher maintains notes for each deployment step. The deployment has five distinct steps, numbered from D1 to D5. These deployment steps are detailed in Chapter 7.5.1. For each deployment step, the researcher notes observations, such as what tool was used, its success, what went well, and what could be improved (see Section 7.5.2). The second data collection, the expert feedback data, is gathered in the one-to-one semi-structured interviews; participants are provided with a demonstration of the deployment tool for data collection. The participants are then asked a series of questions. The researcher gathers notes based on the responses (see Section 7.5.3). In the third data collection, the researcher gathers observations as the ERoPA is utilised to complete the accountability verification process (see Section 7.5.4).

Data Analysis and Findings: The data analysis process for the collected data involves a synthesis to establish the key findings from the case study. The synthesis process starts with identifying the key themes, pattern matching and clustering. These are then prioritised and presented in matrix form to provide a clear, structured way to see relationships and themes within the data, making it easier to spot patterns. Each theme is presented as a row of data in the matrix. These themes are gathered and synthesised against each RoPA system capability (see Section 7.2.3) to establish if ERoPA grants these capabilities. This process is described in more detail in Chapter 7.6.

Validity and Reliability: The rigour and transparency of this research process are upheld through several key steps:

- A series of beta tests were conducted on sample data to evaluate the deployment of ERoPA. The researcher thoroughly reviewed and analysed these tests.
- A pilot study used semi-structured interview questions to refine the data collection methods and improve the questions.
- Samples of the ERoPA video were presented and prepared to determine the most effective content to include and ensure consistency in the vocal track.
- The research integrity was presented to and approved by the Dublin City University Ethics Committee.

7.4 Case Study Design

This section describes the design of the Upsilon ERoPA deployment case study. It represents the technical activities followed to enable the deployment to be studied, such as organisation selection, dataset collection, and tools to support the deployment. This also provides an overview of gathering deployment observational data, expert feedback, accountability verification observational data, and how this information will be presented.

7.4.1 Design for Technical Activities

This section describes the technical approach to enabling the deployment of ERoPA in the Upsilon organisation (see Figure 28 for an overview)

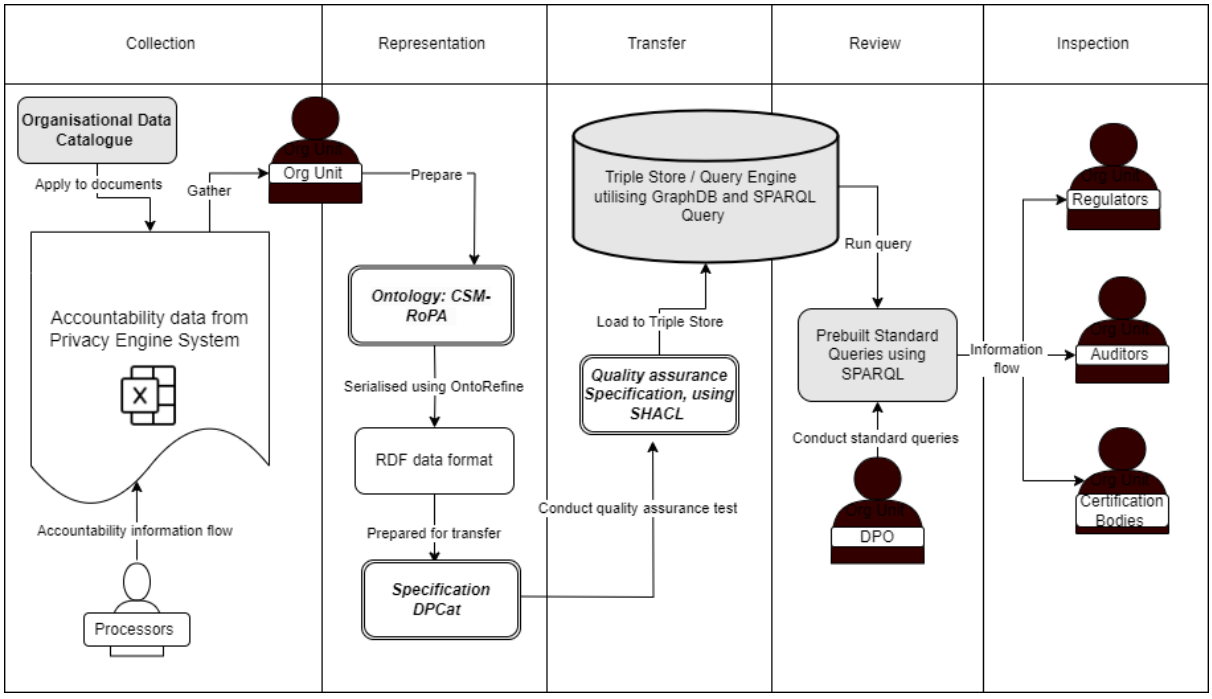

Figure 28 Technical Design for ERoPA Case Study.

Organisation Selection—The ERoPA was deployed in an organisation called 'Upsilon.' The 'Upsilon' organisation is a real organisation, but any information to identify the organisation is removed. Upsilon is a mature, long-established organisation that collaborates with multiple organisational units across many countries. Upsilon engages over one hundred processors to process personal data on their behalf. The organisation processes large amounts of personal data across many categories and subject types. The organisation was selected because it provided a large, diverse, and complex dataset large enough to satisfy the interoperability requirement and complexity necessary to evaluate the ERoPA capability adequately. The Upsilon organisation offers significant complexity and reflects all the stakeholders' data flows that ERoPA needs to represent,

thus providing a suitable deployment case study. The organisation is typical in that GDPR accountability information processed by the case study organisation is held in multiple forms. These storage systems vary from spreadsheets in some organisational units to dedicated privacy systems in others.

The **data set** for the Upsilon deployment was sourced from RoPA records presented in a spreadsheet extracted from a Privacy Engine system. The extract to the spreadsheet is completed using a standard function in Privacy Engine. The dataset contained significant information spread across eleven tabs within the spreadsheet. Each record in each tab has a unique record number, allowing it to be linked back to the relevant data processing activity. Therefore, information regarding processing activity can be associated with its pertinent legal basis, data types, and any applicable international data transfers. The Upsilon dataset contained a consistent vocabulary for describing departments, IT systems, and recipients. This came about because Upsilon used a data catalogue to represent data consistently. This data catalogue approach helps to boost data understanding with detailed information about data assets. This includes where they come from, their quality, who uses them, and how they can or should be used. This information makes it easier for users to grasp the data's meaning, importance, and suitability.

How was the Upsilon data **converted to RDF** and loaded to the triple store? The case study will gather this data in spreadsheet format to load it to the knowledge graph. This data will be converted to a machine-readable format using a schema derived from the Interoperable machine-readable ontology. The Ontorefine tool imports accountability information within GraphDB and converts the accountability information held in the organisational unit to RDF. The ontology uses terms from the CSM-RoPA and the DP-Cat interoperability specifications. These RDF files are loaded into GraphDB[49] repository using the GraphDB import tool[50].

Data validation - The content is subject to a SHACL[51] test to maintain data integrity and quality. The knowledge graph stored in the GraphDB repository stores GDPR accountability information and forms the basis of evidence for demonstrating compliance.

Conduct Typical DPO scenarios: The ERoPA deployment provides several primary use cases for which DPOs might use ERoPA. These standard tasks are identified as part of the Data Protection Professionals survey in Chapter 5.7.2, where the DPOs were asked what they would use ERoPA. The two most frequently occurring tasks identified in the study were (i) Identifying a gap in a RoPA record and (ii) interoperating RoPA with critical stakeholders. In the deployment of ERoPA, the loaded data is stored in the GraphDB, and the knowledge graph can be reviewed using the GraphDB

[49] https://graphdb.ontotext.com/
[50] https://graphdb.ontotext.com/documentation/master/load-your-data.html
[51] https://www.w3.org/TR/shacl/

graph explorer or queried using several prebuilt SPARQL[52] queries to meet the requirements of the DPO. These tasks are described below:

Typical DPO Task 1 is to verify that an Organisational Unit can interoperate with RoPA. In the Data Protection Professionals survey (see Section 4.6.2), it was identified that intra-group and processor interoperability are the primary interoperability requirements of an ERoPA. Typical DPO task one shows how an internal organisational unit (such as HR) can submit RoPA Records using the DPCat specification for inclusion in the ERoPA. This use case seeks to verify that the key stakeholders, such as an organisational unit, can create a RoPA Record and interoperate with the DPO. To simulate this activity in the case study, the Upsilon HR department created and conducted a new business process for the experiment, identified as 26-1. This record was created using the DPCat specification and transformed to RDF, and the record was submitted to the knowledge graph.

Typical DPO Task 2 is to verify that an external data processor can interoperate with RoPA. In typical DPO task two shows how an external entity (such as a processor) can submit RoPA Records using the DPCat specification to include the ERoPA. This use case seeks to verify that that a Processor can submit a RoPA record to the data controller. In task two the processor provided a DPCat record to the controller, which was loaded into the knowledge graph. In this case the record was record number 26-1 which was the same processing activity as task one. (The controller and processor versions of the record are compared in task 3). In practice, the GDPR accountability data supplied by the processor could be gathered from the Data Processing Agreement (see Art. 28), which needs to be in place when a Controller engages a data processor, or the Processor may submit their own Processor RoPA record. `

Typical DPO Task 3 is to identify non-conformance between processor activities and the Controller RoPA. This task identifies if there is any misalignment or non-conformance between the controller record of a processing activity versus the processor's record of the same activity. This is an important task as the processor should only process data on the instruction of the controller (see Art.28). Hence, any misalignment between the processor and controller would be a non-conformance with the GDPR. In Tasks 1 and 2, a processing activity known as 26-1 is loaded into the knowledge graph. The RoPA records are compared using a SPARQL (see Listing 4) to check for 'automated decision-making,' for example. This query could be modified or extended for any of the GDPR concepts gathered for ERoPA (see Section 6.3 for the DPCat specification).

[52] https://www.w3.org/TR/sparql11-query/

Listing 4 SPARQL Query Output Identifying a Conflict on a RoPA record.

```
SELECT DISTINCT ?Record ?org ?AutomatedDecisionMaking
 WHERE {
 ?Record a dpcat:RoPARecord.
 ?Record dct:publisher ?org .
 BIND(EXISTS {
  ?Record dpv:hasContext dpv:AutomatedDecisionMaking
  }as?AutomatedDecisionMaking) .
}
```

Typical DPO Task 4 is to identify a gap in a RoPA record: The GDPR requires every processing activity to have a legal basis, which must be documented in the RoPA based on guidelines. This query helps to identify non-compliant records, enabling retrieval of context, such as which processing activity, who is to be contacted, and what data is involved. This automated process for identifying gaps and conflicts is the primary use case for ERoPA by DPOs identified in the Data Protection Professionals survey (see Section 5.7.2). In the case study, a RoPA processing activity record is created in RDF in the DPCat specification and loaded to Upsilon ERoPA knowledge graph for the experiment. This record intentionally does not contain a legal basis (an attribute that should be maintained on RoPAs) [11]. This activity aims to ascertain if the ERoPA can assist the DPO in identifying the non-compliant record. A SPARQL query is used to find any RoPA Record that does not contain a legal basis entry (see Listing 5).

Listing 5 SPARQL Query to identify Records without a Legal Basis Entry

```
SELECTDISTINCT?Record?date?description?title
WHERE
{
?Recordadpcat:RoPARecord.
?RoPARecordterms:created?date.
?RoPARecordterms:description?description.
?RoPARecordterms:title?title.
```

7.4.2 Design for Observational Feedback on Deployment

Gathering observations during the case study deployment is crucial for gaining in-depth insights into real-life contexts. Such observational data can provide rich, qualitative insights that support or enhance other data types, such as interviews or document analysis. Observations in case studies allow for an authentic view of the context and behaviours, adding a depth of insight beyond what interviews or surveys alone can provide. By carefully planning and structuring observations, one can ensure relevant and reliable data collection that enriches the overall case study analysis [29]. The process used for gathering observational feedback through the deployment following Morgan's framework is as follows:

- Define the focus of the observation – identify the critical activities of the deployment system.
- Observation Method: Conduct an unstructured observation of the deployment activity or procedures that are clearly defined. (Unstructured observation allows flexibility when exploring new environments or phenomena that may not have predefined categories.)
- Researcher Interaction—The researcher conducts a dual role as researcher and supports the deployment, thus gaining a deeper understanding of the context by experiencing the deployment first-hand.
- Schedule observations for each deployment stage so that data is gathered at each process step.
- Maintain notes to ensure that a systematic documentation approach is followed.
- Remain objective and minimise bias - Aim for objective descriptions without interpretation during the observation.
- Capture contextual information that may be relevant and influence outcomes.
- Analyse and synthesise observations regularly to identify patterns for subsequent observations.

Once the deployment observations were gathered, they were synthesised based upon common themes identified and were presented as a summarised set of observations.

7.4.3 Design for Expert Feedback

The case study deployment involved the researcher gathering observational data. This observational data was further supplemented by semi-structured interviews with data protection experts using a prepared video that provided an overview of the deployment. These semi-structured interviews gathered insights from subject matter experts, thus giving a broader perspective into the extent to which the ERoPA supports GDPR compliance.

Table 47 Question Set Utilised in Semi-Structured Interviews.

Question Number	Question from Semi-Structured Interview
1	Would the ERoPA Approach of exchanging RoPA records between stakeholders help DPOs identify gaps, conflicts, and non-compliance?
2	Would the ERoPA Approach of exchanging RoPA records between stakeholders lead to a more comprehensive, up-to-date, and accurate RoPA?
3	How well does the opportunity to round trip the ERoPA with stakeholders overcome some of the buy-in challenges?
4	Would the automated gathering of an Interoperable ERoPA Approach (where data can be gathered from multiple sources such as Data Processing agreements, privacy notices, and other GDPR accountability documents) lead to a more comprehensive, up-to-date, and accurate RoPA?
5	Is this approach better than conventional approaches to RoPA? If so, why?
6	Where do you see any additional uses for this solution within your organisation aside from Article 30 RoPA itself?
7	Are there more critical priorities within the Domain of DPOs that this does not address?

Why were semi-structured interviews chosen? Semi-structured interviews were selected for their effective balance of flexibility and focus, making them ideal for collecting in-depth information, particularly when personal insights are essential. This method blends predefined questions with the opportunity to delve into topics based on participant responses, allowing interviewers to probe deeper into noteworthy points and seek clarification. This interview format encouraged participants to articulate their thoughts in their own words, resulting in richer data that might be overlooked with closed questions. The conversational style fosters a comfortable environment that promotes openness. Core questions facilitate easier comparisons across participants, supporting efficient pattern analysis. The semi-structured interview question set is presented in Table 47.

Video Preparation: A PowerPoint slide presentation with a voice-over narration is ready to provide an overview of the Upsilon ERoPA deployment (see online resource for a copy of the presentation[53]). Using a standard presentation ensures that the same consistent presentation is provided to all interviewees. The presentation contains seventeen slides detailing the purposes of the research, an overview of the Data Protection Professionals (see Section 4.6.2) and the critical issues identified from the study. The presentation provides background information on ERoPA and machine-readable documents. The experts are shown the deployment steps. A series of typical DPO tasks is demonstrated, such as viewing a processing activity, examining a data recipient of personal data, checking technical and organisational measures, establishing the legal basis of a processing activity, and outputting a RoPA in a regulator template format for review. A voice-over narration supported these slides using plain English to explain each slide.

[53] https://doi.org/10.5281/zenodo.14914848

Recruitment of Expert Interview Candidates: The researcher recruited five data protection experts for the semi-structured interviews. The decision to use 'experts' was that the opinions of Data Protection Professionals had been gathered in the earlier survey (see Section 4.6.2). In contrast, the semi-structured interviews would benefit from a smaller, more experienced group of professionals, thus providing greater insight when discussing the implementation of ERoPA. The classification of an expert for this research was that the person had to be or have been a DPO, possess a DPO qualification, and have at least five years of data protection experience. The participants were recruited from the author's public and private enterprise contacts. Two interviewees came from public organisations, and three from private enterprises.

Conduct of interviews: Each Data Protection Expert was invited to attend an online interview. The attendees received a copy of the ethics/consent documentation before the interview commenced. The experts were shown a PowerPoint presentation containing seventeen slides[54]. Once the video was completed, the interviewees were asked the structured questions. These questions were designed to prompt discussion around the implementation and the capabilities of ERoPA to collect, represent, and transfer GDPR accountability information. The interviewer noted the responses from each interviewee. The open nature of semi-structured interviews allowed for additional exploration of responses where appropriate. This allowed the interviews to develop beyond the structured question to explore themes relevant to ERoPA deployment. The interview notes were recorded and stored on a secure Google Drive.

Analysis of results: The results gathered from the expert interviews will be analysed using pattern analysis. This method was chosen because it is valuable for analysing qualitative data due to its flexibility and depth. Pattern analysis offers several advantages, such as uncovering patterns in data, which provides a deeper understanding of complex human experiences. Pattern analysis is straightforward and suitable for researchers at various experience levels. Additionally, it encourages researcher reflexivity, allowing researchers to actively interpret themes while remaining mindful of potential biases in data interpretation. Pattern analysis is suitable for small and large datasets, enabling detailed exploration of each theme and offering nuanced insights into participants' experiences. This transparent method provides a clear path from data to findings, aiding validation. Overall, pattern analysis is a practical method for qualitative research that emphasises depth, flexibility, and transparency.

7.4.4 Design for ERoPA Accountability Verification

In this case study section, ERoPA is used to support the demonstration of GDPR accountability. This work builds on an original publication, 'Demonstrating GDPR compliance with CSM-RoPA' [33],

[54] https://doi.org/10.5281/zenodo.14914848

using the 2021 version of the ICO accountability framework (see Section 3.2). This thesis updates this work based on the ICO accountability framework 2024 version, using ERoPA to verify GDPR accountability.

The approach taken for this section has two parts: (i) the semantic representation of the GDPR concepts within the ICO Framework, and (ii) an analysis of the extent to which the Upsilon ERoPA can meet the stated expectations posed in the framework.

The methodology for this first section of accountability verification (the semantic representation of the GDPR concepts within the ICO Framework) involves the following steps:

1. Analyse the ICO Accountability Framework for the relevant RoPA sections containing GDPR accountability terms contains an extract of Section 6 of the accountability framework (see online resource[55]).
2. Identify all GDPR accountability terms present in the Accountability Framework (see online resource[56]).
3. Deduplicate GDPR terms and identify all unique GDPR terms. To derive these, the researcher performed term extraction, semantic analysis, term frequency enumeration, de-duplication, and antonym/homonym identification. The criteria for including or excluding terms are based on whether their characteristics are defined within the GDPR text. If a term is not defined within the GDPR text, it is excluded from the list of information categories (see online resource[57]).
4. Compare the unique terms to ERoPA to establish if they have a corresponding exact pattern match of each other, a complex match, a partial match, a match with another vocabulary or no match [24]. This will support the analysis of complexity/expressivity of term (see Appendix I).
5. Gather observations through the deployment of ERoPA for accountability verification.

The methodology for this second section of the accountability verification (meeting the regulator's expectations set out in the ICO Framework) involves the following steps:

1. Identify the expectations within the ICO framework relevant to RoPA
2. Examine these expectations to assess the extent to which the GDPR information gathered in the Upsilon Case Study can meet these expectations.
3. Provide a link to the case study for the assessment

[55] https://doi.org/10.5281/zenodo.14914848
[56] https://doi.org/10.5281/zenodo.14914848
[57] https://doi.org/10.5281/zenodo.14914848

4. For each expectation, establish the extent to which the Case Study ERoPA deployment can be used to represent each expectation. This information is presented as fully, partially, or not met.
5. Gather observations through the deployment of ERoPA for accountability verification.

7.4.5 Presentation of Findings

The previous three sections describe the data collection process in the Upsilon case study deployment. This data comes as observation notes collected during the deployment, notes gathered as part of the expert interviews, and observations collected as part of the ERoPA accountability verification. This section describes how these three data sources are analysed and synthesised to provide the case study's findings. These findings enable the evaluation of the extent to which the application of the ERoPA supports the requirements for deploying GDPR accountability and helps to identify critical guidelines for organisations deploying ERoPA in practical use cases.

Pattern Analysis and Matching: As three distinct data sources were collected, this research looks for common themes or patterns across different data sources that aid in the clustering of similar data points. The research identifies consistent themes or patterns and groups this data into clusters with similar characteristics. This pattern-matching process is used to identify themes and relationships within the qualitative data collected as part of the case study. This process is completed by comparing empirical patterns found in the data to theoretical or expected patterns to derive potentially generalisable insights. This analysis involves identifying qualitative data to identify recurring themes, which can serve as the basis for developing implementation guidelines. The pattern matching is completed across three sources of data, thus leading to a more robust set of findings, and avoiding unique organisational, cultural, or operational contexts observed in a single data source case study. This multiple-source triangulation approach supports the **Reliability** of the qualitative conclusions [155]. This rigour helps bolster the credibility of implementation guidelines derived from case study data as the research process becomes more transparent and replicable.

Prioritisation and presentation: To derive actionable insights from the Upsilon ERoPA deployment case study, the three data collections: (i) deployment observations, (ii) the expert feedback, and (iii) the ICO accountability framework verification and gathered and synthesised against each RoPA system capability (see Section 7.2.3) to establish if ERoPA grants these capabilities and to establish any additional missing capabilities.

7.5 Data Collection

This section describes the data collected through each stage of the technical deployment and the data gathered from the Expert interviews. The five steps for each technical deployment step are described in Chapter 7.5.1. Observations are collected at each one of these deployment steps. The video was prepared upon completion of the deployment, and the semi-structured interviews were conducted. A summary of the data collected in these interviews is presented in Chapter 7.5.2.

7.5.1 Deployment Process

D1-Dataset Preparation: The Upsilon data set came from a Privacy Engine commercial privacy management software system[58]. The data was exported from the privacy system. The format of this information was that it was presented as a spreadsheet document containing eleven tabs of information. The tabs represented different data element types, such as a tab containing IT systems and another tab containing third-party entities (see Table 48 for a detailed description of the dataset and the content of each tab). Each of these tabs had a common reference number for each processing activity relating to a unique activity, thus enabling the assembly of a model for an end-to-end processing activity using the unique reference number.

Table 48 Description of Dataset Used for Case Study.

Spreadsheet Tab Name	**No of Records**	**Description of Record Types and Categories in this Tab**
Data Processing Activities	246	3 Distinct companies with 246 entries entered in a hierarchical structure
Intra Organisational departments	14	Organisational units: Human Resources, Group Finance, Facilities, Quality, Corporate, Finance, Commercial, Customer Service, Operations, Legal, IT Services, Finance, Distributors, Distributors Finance.
Third-Party entities	97	75 Data Processors, 21 Data Controllers, 1 Data Recipient
Third-Party documents	141	Non-disclosure agreements, Data Processing agreements, Contracts (unstructured data)
Transfers Abroad	9	India, United Kingdom, United States of America
Data Types (processed)	10	Location data, any direct or indirect reference to a living individual which identifies that person, Financial Information, Government identifiers, Data relating to health or sexual life and sexual orientation, non-sensitive personal information, Trade union membership, Racial or ethnic origin,

[58] Data Extract was gathered on 17/02/2023

Spreadsheet Tab Name	No of Records	Description of Record Types and Categories in this Tab
		Data of children aged sixteen or under, Religious or philosophical beliefs.
Data Subject Categories	10	Current Employees, Customers, Former Employees, Job Candidates, Other Suppliers, Vendors, Visitors, Website B2B visitors, and Website B2C Customers.
System	58	Named IT systems
Acquisition Method	9	CCTV, Email, Geo-tagging, HR System, In -Person, Network, Logs, Online Form, Paper form, Phone
Legal Basis	10	Assessment of working capacity, Compliance with legal obligation, Consent, Contractual necessity, Employment law, Explicit consent, Legal claims, Legitimate interests, medical diagnosis, Preventive or Occupational medicine
Data Elements	43	Bank / Financial Services Account Number, Call Recording, CCTV data, Communication Data, Contact Data, CV, Date of Birth, Disability / Mobility Details, Driving licence, Email address, email content, Employee Benefits, Employee Role Data, Employment History, Expenses Data Financial and Tax Data, Health Data, Home address, Identification data, Landline number, Medical Condition / Diagnosis, Meeting video recordings, Mobile phone number, Name, Nationality, Next of Kin, Passport Number, Performance Data, Postal Address, PPS / Social Security Number, Qualifications, Racial / Ethnic Identity, Recruitment Data, Referee Data, Religious Affiliation / Denomination Sexual Orientation / Designation, Special Category Data, System Logs, Title, Training Records, Travel Data, Unique Customer Identifier, Unique Employee Identifier

The data is primarily semi-structured, as controlled data terms and concepts describe data processing activities. The only exception is ‘third-party documents, ‘ largely unstructured data.

D2-Conversion to RDF: Converting the spreadsheet data extracted from Privacy Engine into RDF requires mapping rules to be created for each of the eleven tabs of the spreadsheet. These rules were constructed using the OntoRefine tool [185], part of the GraphDB Free software product built on the open-source OpenRefine tool [157]. OntoRefine is a data transformation and enrichment tool often used in semantic data projects, particularly for preparing data with knowledge graphs and other linked data systems. OntoRefine provides mapping rules for semantic ontology alignment to support RDF (Resource Description Framework) transformations. The columns in each spreadsheet are mapped to `classes and properties from the CSM-RoPA and DPCat ontologies. An example of the mapping from Privacy Engine to RDF is presented in Table 49.

Table 49 Example of Ontorefine Mapping to ERoPA

dct:identifier	dpv:hasPersonalDataHandling	dpv:hasResponsibleEntity
Activity ID	Activity Name	Department
24998	Onboarding new employees (Offer)	Human Resources
25000	Finance Activities	Human Resources
25001	Payables processing	Finance Upsilon
25003	Physical Security	Facilities
25007	Quality Compliance	Quality

Once the mappings are constructed, the Ontorefine mapping interface can be used to output the contents of the spreadsheet tab as an RDF file. Appendix J presents a sample of an RDF file generated using OntoRefine.

D3- Load to the Upsilon Data Protection Knowledge Graph: The mapped data is converted to RDF and loaded to the Upsilon ERoPA knowledge graph using the GraphDB workbench. The mapped data loaded data was explored on GraphDB Ontorefine using the explore function. A sample RoPA record visualisation fragment is displayed in Figure 29. This visualisation shows the GDPR processing information (such as the data source and the data category) related to a specific RoPA record. The colour of each node in the diagram represents different node types to support the visual differentiation of each type of node.

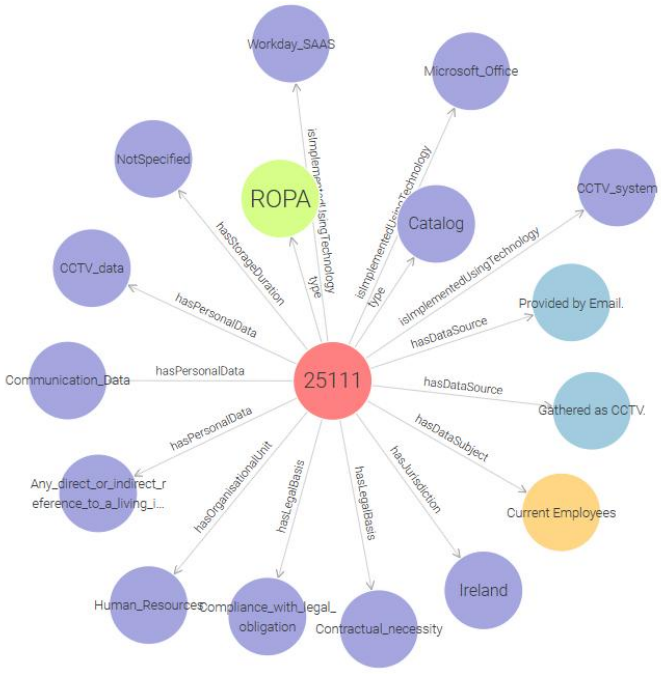

Figure 29 Visualisation of a RoPA Record Fragment an ERoPA Knowledge Graph[59].

D4-Validation: Two methods were used for conformance and provenance verification for the Upsilon Case study. Firstly, the DPCat specification was used when information was collected and transferred to the Upsilon ERoPA knowledge graph. The DPCat specification details the information

[59] Source: Upsilon deployment of ERoPA

required for RoPA and RoPA Records. This cardinality information requirements are set out in Table 33 for RoPA and Table 34 for RoPA record (see Section 6.4). The second method used SHACL shapes to ensure that RDF graphs meet predefined conformance and provenance requirements. The specifications for the SHACL shapes for ERoPA are available online[60].

For the Upsilon case study, the 246 processing records were converted from spreadsheet RoPA data into RDF format and checked for conformance and provenance requirements. In addition, a manual RoPA record was created using the DPCat Specification. These RoPA records loaded successfully, meeting the necessary provenance and conformance requirements.

D5- Conduct of Typical DPO tasks: This section describes four typical DPO tasks conducted during the deployment of ERoPA. The design for technical activities fully describes these tasks (see Section 7.4.1).

Typical DPO Task 1 was to verify that an Organisational Unit can interoperate with RoPA. This task seeks to confirm that key stakeholders, such as an organisational unit, can create a RoPA Record and interoperate with the DPO. To simulate this activity in the case study, the Upsilon HR department created a new business process for the experiment. This process was identified as 26-1. This record was made in RDF format using the DPCat specification and successfully loaded to the knowledge graph. There are distinct differences when one compares this to an equivalent Privacy Engine RoPA data load. Privacy Engine, like many vendor-supplied privacy software systems, offers an upload tool to import RoPA. However, the ongoing maintenance of RoPA records requires manual input to the Privacy Engine system, often by the DPO but sometimes by nominated data champions representing organisational units. The key difference between ERoPA and the existing privacy system is that automation is used to load the RoPA Record.

Typical DPO Task 2 is to verify that an external data processor can interoperate with RoPA. This task shows how an external entity can submit RoPA Records using the DPCat specification to the data controller ERoPA. To simulate this activity in the case study, a simulated third-party data processor called SAP provided a DPCat record to the controller (see Appendix K). This data RoPA record was manually prepared and loaded into the knowledge graph. The record loaded successfully to the ERoPA.

In practice, the provision of such processing data activity records by a processor would be a manual activity. In general, this information would be reviewed before processing commenced and at agreed intervals on processing has commenced as part of annual due not be presented as a RoPA record but would form part of a Data Processing Agreement (DPA) (see Art.28), a Data Protection Impact Assessment (DPIA), or supplier due diligence process. The processor would also

[60] https://github.com/Paul-Ryan76/DPCat

maintain its own processor RoPA containing the details of the processing activity on behalf of the controller. Hence, the validation of the processing record may need to be reconciled against DPA or other records, such as the processor's own RoPA. This process contrasts significantly with the Privacy Engine system, where the data processing agreement is stored as a document, such as a PDF, which can be attached to the processing activity record. However, the automated checking of the data processing agreement (or other compliance documents) against the processing record is not possible.

Typical DPO Task 3 is to identify non-conformance between processor activities and the Controller RoPA. For typical DPO task three, the processor RoPA Record was added to the knowledge graph in the experiment. The Controller and Processor records were then compared using a SPARQL query (see Listing 4), and a conflict was identified in record 26-1. The output of the SPARQL query is displayed in Table 50. The record shows that the controller RoPA was not aligned with the processor record of the actual processing that was occurring[61].

Table 50 Sample of SPARQL Report used to Identify Conflict between RoPA Records.

Ropa Record Number	**Organisation**	**Automated Decision Making**
https://w3id.org/DPCat/examples/Upsilon/26#26-1	UpsilonPLC	No
https://w3id.org/DPCat/examples/Upsilon/26#26-1	SAP	**Yes**

Typical DPO task four is to identify a gap in a RoPA record. The GDPR requires every processing activity to have a legal basis, which must be documented in the RoPA based on guidelines. In the case study, a RoPA processing activity record is created in RDF in the DPCat specification and loaded to the Upsilon ERoPA knowledge graph for the experiment. The outcome of the SPARQL query is that RoPA record #23-1 is returned as having no Legal basis entry (see Figure 30).

	Record	date	description	title
1	https://w3id.org/dpcat/examples/Upsilon/23#23-1	"2023-05-15"^^xsd:date	"Customer Service Processing Activities"@en	"Customer Service Procesing Activities "@en

Figure 30 SPARQL Query Output Identifying Non-Compliant RoPA Record

[61] Note: record 26-1 is a manually amended record and in no way should it be seen as an indicator that SAP utilise automated decision making when processing personal data

This use case shows that ERoPA can successfully identify a gap in a RoPA entry. The same SPARQL query process could be repeated for any mandatory RoPA Record entry field to establish whether there are gaps in a RoPA record. The typical DPO tasks show the following:

Task 1: An Intra Organisation RoPA record has been successfully loaded to the ERoPA to demonstrate interoperability between the organisation and the Controller.

Task 2: A Data Processor has successfully provided the controller with a RoPA record, which has been loaded to demonstrate processor interoperability.

Task 3: The comparison of the Processor and the Upsilon PLC HR RoPA record identified a non-compliance.

Task 4: A RoPA record that did not contain a legal basis for processing personal data was identified.

7.5.2 Observations made through the deployment process

The Upsilon ERoPA deployment had five steps (described above). Observations were noted and gathered during each step, which are presented in Table 51.

Table 51 Observations Gathered through the Upsilon ERoPA Case Study.

Deployment stage	Ref. No.	Observation
D1 Gather data	1	The data extracted from the existing privacy Engine system was presented in a structured spreadsheet containing eleven tabs, making the mapping process easy.
	2	The Upsilon organisation did not have any form of data catalog or structured data governance, and it did not have any metadata management capability. Whilst it was possible to extract the Privacy Engine data and convert it to RDF using a Schema, the unstructured nature of data in the Upsilon organisation could create significant effort when converting GDPR-relevant information that resided outside Privacy Engine to RDF.
	3	Each record extracted from the Privacy Engine system had a unique identifier number, allowing the RoPA record to be assembled from the relevant fragments in each tab. The unique identifier made the process of combining each of the eleven tabs of data easier. The complexity of larger data sets from diverse sources may create more significant challenges.
D2. Convert to RDF	4	In the case study, the Ontorefine transformation tool was easy to use and met all the needs to convert the data to RDF. This step was necessary to convert the data from the Privacy Engine spreadsheet into RDF. A RoPA record was also loaded using a manually written RDF file. The mapping process for Ontorefine transformation and the manual creation of an RDF file required some knowledge of the DPV terms used in DPCat, but once the user had some knowledge of RDF and DPV, the activities were easy to complete.

	5	The CSM-RoPA ontology and DPCat Specification contained all the terms needed to prepare the mappings and represent the data in the Privacy Engine system.
	6	The mappings created for the conversion were based on how the Privacy Engine software exported the data to a spreadsheet. If the exported data structure changes, the mappings may need to be modified. This would require personnel with technical knowledge to complete a partial re-mapping.
D3. Load to Triple Store	7	The GraphDB upload tools allowed uploads from URLs, files, or snippets. This proved to be user-friendly and provided error messages for any load errors.
	8	The RDF files were loaded to GraphDB successfully. However, some files required several attempts due to mapping errors. This required a remapping by the practitioner to resolve mapping issues.
	9	the GraphDB explorer functionality worked well in visually checking and sampling the Upsilon knowledge graph. The explore functionality was utilised to search for sample records by record number, and these were checked against the Privacy Engine system and the spreadsheet.
D4. Validate	10	SHACL constraints proved useful for ensuring information correctness and checking whether the necessary information is present and has expected values.
	11	SHACL constraints proved useful for evaluation, checking whether the necessary information is present and has expected values. Based on requirements drawn from the GDPR, such as ensuring the purpose of processing is declared, they thus directly address GDPR compliance verification.
	12	Some hurdles were faced in representing the RoPA information using DPV, as the DPV 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.
D5 Conduct DPO tasks	13	The standard DPO tasks, such as finding RoPA gaps or conflicts, worked well.
	14	The second use case, where a processor submitted a RoPA Record for inclusion in the knowledge graph, is a simplified process. However, this may be harder to achieve in practice, as the RoPA data may need to be gathered so stakeholders can interoperate with RoPA and exchange data successfully.
	15	Task one involved gathering a RoPA record from an organisational unit. Whilst this task was completed, it did not overcome one of the challenges of gaining buy-in from the organisation units [11]
	16	Task four involved the identification of a RoPA record that did not have a legal basis in place. While this task was completed, there was a need to specify the queries that the DPO required or to develop a GUI to support the DPO in this task. Also, it should be noted that the Privacy engine RoPA feature is already present in that Privacy Engine where such gaps are missing a legal basis.

7.5.3 Semi-structured Interview Data

The interviews with data protection experts were conducted between November and December 2024. The following are summarised notes, critical statements, and interview comments. The full interview notes can be found in Appendix L.

Expert one:

- The ERoPA Approach's benefit is that it eliminates a lot of manual work
- Using ERoPA for validation versus other documents is good, particularly if you can triangulate.
- The advantage of the ERoPA Approach is that it will lead to a more granular ROPA.

Expert Two:

- RoPA is a 'high-level crib sheet '. It does not have all the details of all GDPR compliance data. It needs to be associated with other compliance documents.
- Exchanging RoPA information with processors and organisational units is a good idea. This should lead to better accuracy and more up-to-date ROPA.
- The idea of scheduling ROPA reviews automatically is good.
- When comparing your ROPA with a data processing agreement or a privacy notice, you must be careful, as the language is different (depending on the audience). You may not be comparing like with like.
- If you are to machine-read documents, you need to have agreed and defined semantics.
- Regulator support for the implementation of ERoPA would be beneficial.

Expert Three:

- Everyone expresses challenges when dealing with an Excel-based RoPA.
- So many contact points require a contribution to ROPA. DPOs will like this approach to RoPA.
- Using Excel for RoPA is complex, challenging, and time-consuming.
- Agree that ERoPA can support the identification of compliance gaps.
- The use of standard semantics in ERoPA is essential to help compliance, but one needs to be careful of the semantics of the document's meaning.

Expert Four:

- The interoperability of ROPA is a great idea, but it needs common agreement on terms.
- The regulator's support in getting acceptance of ERoPA would be good.
- Knowledge graphs in ROPA are very good, as the RoPA references many other documents.
- The interoperability of information for RoPA is critical.

- Gathering RoPA information from the relevant personnel in a form suitable for RoPA can be challenging.
- Aligning privacy notices, data processing agreements, and ROPA is a long way off within organisations.
- The deep dive of RoPAs by regulators has confirmed that ROPAs are too shallow and need more detail.

Expert Five:

- A lot of RoPAs are Excel-based – any automation would help.
- It can be difficult to check different documents for GDPR compliance, and there may be a lack of consistency across them.
- The interoperability of ROPA between stakeholders is a great idea, but common semantics are essential for interoperability.
- Exchanging RoPA information between stakeholders is a good approach for compliance verification. Using other documents to verify ROPA would be excellent and work well.
- Exchanging RoPA may not overcome the lack of buy-in among stakeholders; however, it may work well with processors as buy-in can be forced on vendors/processors.
- The ERoPA interoperability offers advantages for processors as they can align compliance with the controller and reduce expensive audits and manual data exchange and review.
- The critical thing with intra-organisation compliance verification is to make it easy for the stakeholders.

7.5.4 ERoPA Accountability Verification Data

This section describes the data collected from ICO accountability framework verification. This data comes in two parts: (i) data gathered from the representation of the ICO accountability framework using DPCat and (ii) an assessment of the extent to which ERoPA can support ICO accountability framework verification.

Representation of the ICO Accountability Framework with DPCat: The process identifies each accountability expectation's unique terms (concepts and relations) in section 5.2. The mapping methodology used the same method (as CSM-RoPA) described in Section 5.2. It was conducted on the ten expectations (sections) and the thirty-three questions in the 'Records of processing and the lawful basis' section. The mapping steps are displayed in Figure 31

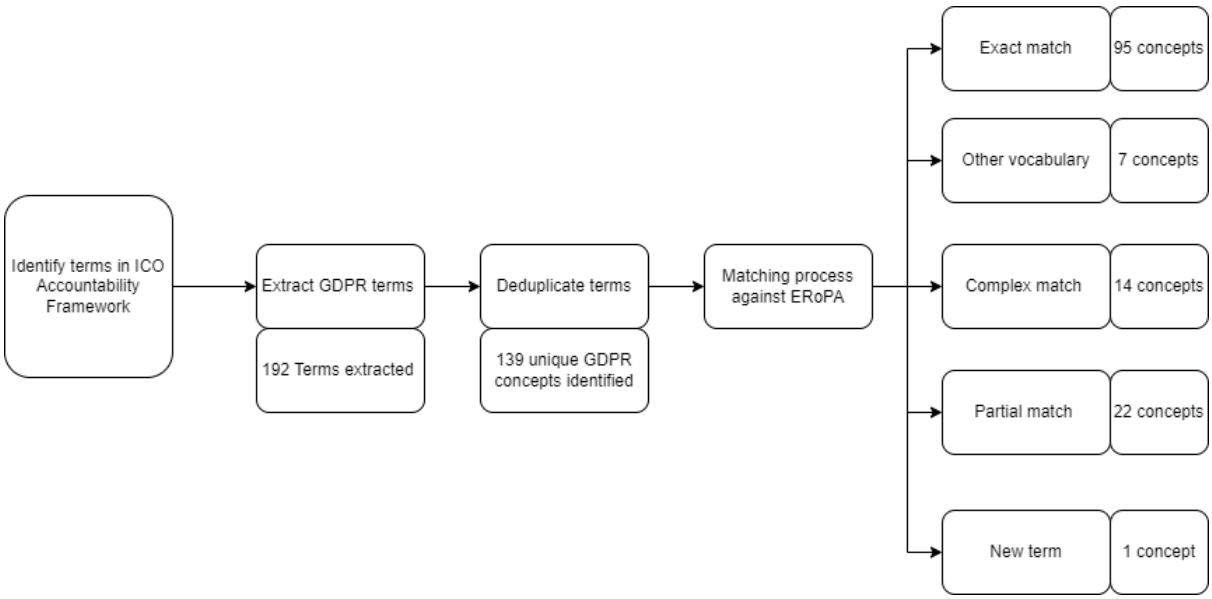

Figure 31: Extent of DPCat coverage to represent the ICO Accountability Framework

A summary of the result of the mapping accountability verification exercise is presented in Table 52 below.

Table 52 ICO terms to ERoPA Mapping.

Match Expressivity	**No of terms**	**% of all terms**
Exact Match	95	68.3%
Other vocabulary	7	5.0%
Complex match	14	10.1%
Partial match	22	15.8%
DPV Under consideration	1	0.7%
Grand Total	139	

Utilising ERoPA to meet the ICO Accountability Framework: This second section of the accountability verification process examines each expectation within the ICO framework to establish the extent to which GDPR information gathered in the Upsilon Case Study can meet these expectations. Table 53 provides an overview of each expectation and how it is met in the Upsilon case study. For each expectation, the table provides examples of how the expectation is met and the extent to which it is met.

Table 53 Extent to which the Upsilon ERoPA meets Regulator Expectations.

Regulator Expectation	Reference	Ways to meet the regulator's expectations	How is this achieved with the Upsilon ERoPA	Achieved
Data-mapping: Your organisation frequently conducts comprehensive data mapping exercises, providing a clear understanding of what information is held and where.	6.1.1	Your organisation conducts Information audits (or data mapping exercises) to find out what personal data is held and to understand how the information flows through your organisation.	In the case study all data from the Privacy Engine system was gathered and converted to RDF using the DPCat specification. While the representation of Privacy Engine supplied data was achieved, there may be additional GDPR information held in other systems and documents, which requires further collections and representation.	Partial
	6.1.2	The data map is kept up-to-date, and you assign the responsibilities for maintaining and amending it.	The DPCat records are loaded to the triple store in Upsilon. Each process activity has a publisher and temporal data, thus supporting provenance and conformance. The Upsilon RoPA contains 3 Distinct companies with 246 RoPA entries, each with a temporal record (see Section 7.5.1).	Fully
	6.1.3	You consult staff across your organisation to ensure an accurate picture of processing activities, for example, by using questionnaires and staff surveys.	In Upsilon, the data processing activities are stored in the Triple store. These records can be output as a DPCat specification or as a report for review by relevant stakeholders. This is referred to as a round trip in the Expert survey (see Section 7.5.3).	Fully
	6.2.2	Your organisation regularly reviews the record against processing activities, policies, and procedures to ensure it remains accurate and up-to-date, and you assign responsibilities for doing this.	The Upsilon processing data is held in a triple store. These records could be compared against the policies and procedures. Ideally, this would be best achieved if the policies were machine-readable.	Partially
	6.2.3	You regularly review the processing activities and data types you process for data minimisation purposes.	The DPCat records are loaded to the triple store in Upsilon. Each processing activity has a publisher and temporal data, thus supporting provenance and conformance. These records can be reviewed by DPO using the query functionality. In terms of data minimisation, the knowledge graph contains detailed information on each business process such as data types, retention period and purpose of processing, however the RoPA alone may not contain sufficient granularity to make this assessment.	Partially

Regulator Expectation	Reference	Ways to meet the regulator's expectations	How is this achieved with the Upsilon ERoPA	Achieved
ROPA requirements: The ROPA contains all the relevant requirements in Article 30 of the UK GDPR.	6.3.1	The ROPA includes (as a minimum): •Your organisation's name and contact details, whether it is a controller or a processor (and where applicable, the joint controller, their representative and the DPO); •the purposes of the processing; •a description of the categories of individuals and personal data; •the categories of recipients of personal data; •details of transfers to third countries, including a record of the transfer mechanism safeguards in place; •retention schedules; and •a description of the technical and organisational security measures in place.	The ERoPA case study utilised SPARQL query to output RoPA in this standard format (see Section 7.5.1).	Fully
	6.3.2	You have an internal record of all processing activities conducted by any processors on behalf of your organisation.	The Upsilon case study recorded seventy-five data processors conducting activities on behalf of the organisation (see Section 7.5.1).	Fully
Good practice for ROPAs: Your organisation's ROPA includes links to other relevant documentation as a matter of good practice.	6.4.1	The ROPA also includes, or links to documentation covering: •information required for privacy notices, such as the lawful basis for the processing and the source of the personal data; •records of consent; •controller-processor contracts; •the location of personal data; • DPIA reports; •records of personal data breaches; •information required for processing special category data or criminal conviction and offence data under the Data Protection Act 2018 (DPA 2018); and •retention and erasure of policy documents.	Upsilon DPCat specification contains the 43 GDPR concepts in regulator RoPA templates. This includes all concepts found in this expectation. An example of this is that there are 141 third-party documents, such as data processing agreements (see Section 7.5.1).	Fully

7.6 Data Analysis

This section gathers the data collected in the Upsilon case study and analyses it to evaluate the feasibility and capability of the ERoPA Approach. This section evaluates the extent to which the application of the ERoPA supports implementing GDPR accountability (RSQ3). The findings from the analysis are used to identify the key considerations for organisations implementing an ERoPA Approach and to develop a structured approach to support organisations in implementing ERoPA (RSQ4).

7.6.1. Analysis of Direct Observations of Deployment

In the upsilon case study deployment, sixteen observations were gathered. These observations are presented in Table 51. This section provides an analysis of the observations to identify the key themes emerging from the observations The key observation themes are listed below and labelled DO1-5, which **DO** stands for. The following themes have emerged from the direct observations.

DO1 Standard ontologies and Specifications - The DPV Vocabulary, the CSM-RoPA ontology and the DPCat Specification contained all the terms needed to prepare the mappings, and the cardinality requirements of DPCat ensured that mandatory dataset fields were met. The DPCat standard specification supported the interoperability of ERoPA information with relevant stakeholders. (Refer to 7.5.2 observations ref no. 1,10,11,12)

DO2 Deployment Supporting Tools—The tools used to conduct the deployment, such as Ontorefine, GraphDB triple store, and SPARQL, met the deployment's needs and were easy to use. The GraphDB explorer functionality worked well to visually check knowledge and enable the viewer to find data quickly. In general, the tools worked well in completing the deployment. SHACL provided a very useful tool for validation and conformity, and the GraphDB load error messages proved helpful. (Refer to 7.5.2 observations ref no. 4,5,7,9,10,11)

DO3 Data Governance - The presence of standard data terms across the data set made the dataset preparation much easier than it might have been. An organisational data catalogue/ governance is particularly beneficial to help structure the data. (Refer to 7.5.2 observations ref no. 2,3)

DO4 Tools to support Typical DPO tasks— ERoPA requires querying tools to support typical DPO tasks. The Upsilon deployment provided two typical DPO tasks, which were successfully achieved. (Refer to 7.5.2 observations ref no. 13,16)

DO5 Modelling approaches—The Schema created for the conversion was based on how the Privacy Engine software exported the data to a spreadsheet. If the exported data structure

changes, the schemas may need to be modified. Also, there may be challenges if the organisation model (using the same concepts) takes differing approaches to modelling. Hence, there may be merit in developing best practices for modelling RoPA. (Refer to 7.5.2 observations ref no. 3,12)

7.6.2 Analysis of Expert Feedback from Semi-structured Interviews

Section Chapter 7.4.3 details the approach to analysing the results gathered from the expert interviews using pattern analysis. The responses from the experts were gathered and processed using pattern analysis. The themes identified from the data protection expert interviews are summarised in Figure 32. The key findings from the expert group provided feedback in several areas.

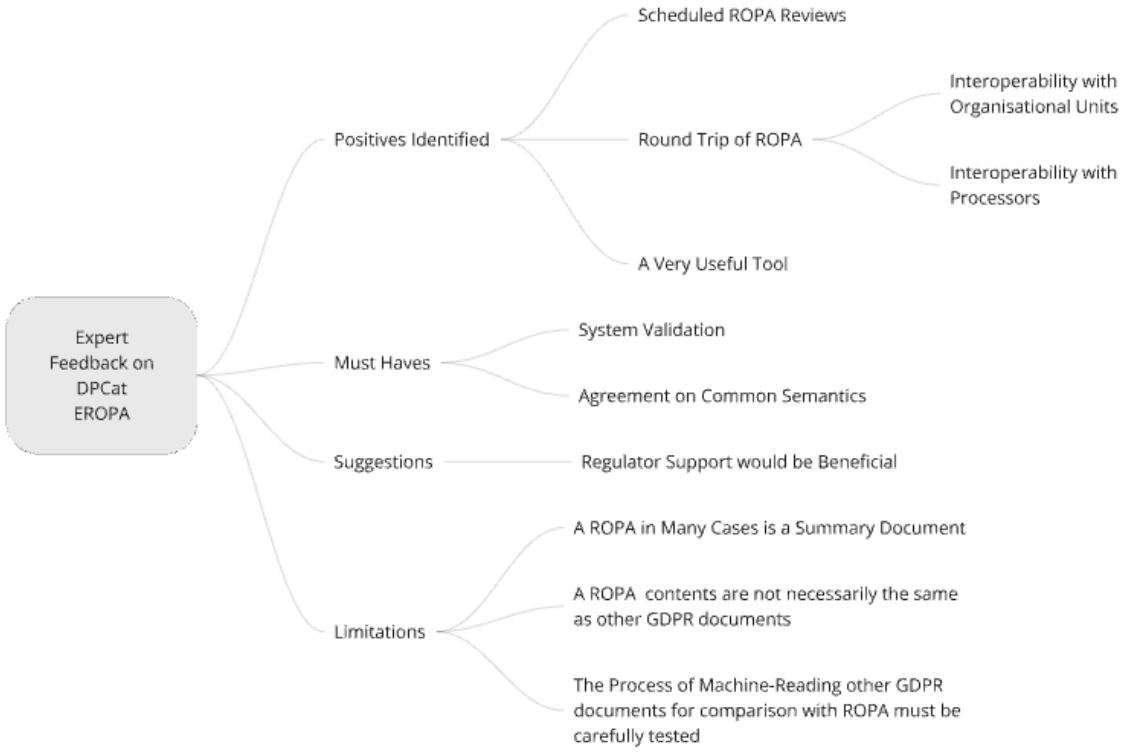

Figure 32 Analysis of Themes from Semi-structured Interviews

The expert group consider the ERoPA Approach an exciting approach to RoPA that could lead to significant efficiencies. The expert group gave positive and negative feedback on ERoPA in six areas (these are labelled EF 1-6 below, where **EF** stands for expert group feedback). The following themes have emerged from the five interviews with experts.

EF1 The ERoPA Approach is very useful to support typical DPO tasks. The experts acknowledged that ERoPA helps identify gaps, inconsistencies, and non-compliance (see Section 4.7 requirement 3.6). ERoPA would be particularly beneficial to DPOs working in a spreadsheet-based domain.

EF2 ERoPA interoperability with stakeholders was a strong positive. The RoPA exchange with the organisational units and data processors was identified as a strong positive for the ERoPA Approach. The experts proposed the concept of a scheduled RoPA exchange as a potential benefit for keeping RoPA up to date. The ERoPA Approach may not overcome organisational buy-in challenges, but it may be more successful in gaining buy-in from data processors due to the contractual relationship between parties.

EF3 ERoPA data transfers require a standard specification: The expert group has identified that ERoPA must operate with agreed standard semantics, which is critical to avoid ambiguity between GDPR concepts in different documents (see Section 4.7 requirement 1.3).

EF4 An enabling Regulator would support the adoption of ERoPA. The experts have suggested that the success of ERoPA Approach would be significantly enhanced if national regulators or the European Data Protection Board (EDPB) supported a systematic automated approach to GDPR compliance. This is considered a critical success factor in RegTech, but it is currently not a feature of GDPR regulatory compliance.

EF5 The importance of an organisational data catalogue in supporting ERoPA. The expert group has identified that although RoPA is an essential compliance document, it does not contain all-encompassing information. In many cases, RoPA includes a summary of processing activities, but it should be considered a reference document that leads to other supporting compliance documents. A catalogue approach enables the DPO to access the granular information necessary to demonstrate compliance.

EF6 ERoPA must be capable of machine-reading other GDPR documents to compare with RoPA. The experts agree that ERoPA is a valuable resource as a stand-alone document. However, to fully benefit from the ERoPA Approach, it must be linked to supporting GDPR documents. This must be tested very carefully. For example, comparing a privacy notice with RoPA, a standard DPO task, may not yet be possible with ERoPA as the language of a RoPA may differ from that of a Privacy Notice, and the same may be true for other GDPR accountability documents. However, a standard vocabulary, such as the DPV, supports this.

7.6.3 Analysis of ICO Accountability Framework Verification using ERoPA

The accountability verification for the case study examined how effectively ERoPA can be used to support the RoPA requirements outlined in the ICO Accountability Framework. The analysis revealed that ERoPA can capture all necessary information concepts to represent the nine RoPA expectations set by regulators. Furthermore, ERoPA includes mechanisms like DPCat to gather and validate this information in realistic scenarios.

The following six themes emerged from this accountability verification analysis, and they are numbered AV1-6, where **AV** stands for accountability verification.

AV1 The stakeholders play an important role in RoPA as an emerging team. Organisational units are key entities that provide information processing to RoPA. Their role in information audits is crucial to provide the DPO with up-to-date and accurate information. The organisation unit must supply/ review RoPA records to ensure an up-to-date and accurate RoPA. The regulator identifies explicitly the importance of the activities conducted by processors as an expectation. In the Upsilon deployment, the activities conducted by these seventy-five processors were recorded on the RoPA. (see 7.5.4 refer to expectation 6.1.1, 6.1.2,6.1.3, 6.2.3 and 6.3.2 of the ICO framework)

AV2 The Information held in ROPA alone is not sufficient to demonstrate accountability. While the representation of Privacy Engine extracted spreadsheet data was successfully loaded to ERoPA, other GDPR information may be held in other sources that would be required for accountability verification. Examples of this would be Data Processing Agreements, Data Protection Impact Assessments, privacy notices, policies, and procedures. Whilst the Privacy Engine System identified these as compliance documents, there would be significant benefit if these documents were machine-readable to support accountability verification (see 7.5.4 refer to expectation 6.2.2, 6.2.3 and 6.4.1 of the ICO framework).

AV3 Provenance and conformance of RoPA records. The regulator identifies that RoPA stakeholders should regularly review the record against processing activities, policies, and procedures to ensure it remains accurate and up-to-date and assigns responsibilities for doing this. The DPCat specification requires conformance data to be provided to support provenance. The exchange of RoPA records between stakeholders enables stakeholders to maintain up-to-date and accurate records whilst ensuring that mandatory conformance requirements are met (see 7.5.4 refer to expectation 6.2.2, 6.2.3 of the ICO framework).

AV4 Regular reviews of processing activities. RoPA Records must be reviewed regularly to ensure they are up-to-date and accurate. The importance of query tools for the DPO to export a DPCat record for review by an organisational unit or expectation reporting where a RoPA record may require attention. (see 7.5.4 refer to expectation 6.2.2, 6.2.3 of the ICO framework).

AV5 ERoPA provides a comprehensive Semantic coverage of GDPR concepts for RoPA. The analysis identified that 98.6% of the concepts required for section 6 could be represented using ERoPA. The foundation ontology CSM-RoPA becomes using DPV concepts supported by the DPCat interoperability specification can support the collection, representation, and transfer of GDPR information (see 7.5.4; refer to expectations 6.3.1, 6.3.2, and 6.4.1 of the ICO framework).

AV6 Supporting Typical DPO tasks. For Upsilon to meet the expectations of the ICO accountability framework, the DPO utilised ERoPA components to support typical DPO activities.

Among these are reporting tools such as SPARQL query (see 7.5.4 refer to expectation 6.2.2, 6.2.3, 6.3.1, 6.3.2 and 6.4.1 of the ICO framework).

7.6.4 Case Study Synthesis

This section derives actionable insights from the Upsilon ERoPA deployment case study. The three data collections: (i) deployment observations, (ii) the expert feedback, and (iii) the ICO accountability framework verification and gathered and synthesised against each RoPA system capability (see Section 7.2.3) to establish if ERoPA grants these capabilities and to establish any additional missing capabilities. This information is presented in Table 54.

Table 54 Synthesis of Case Study Data Analyses

ERoPA capability	**Emerging Theme from Analysis**	**Data Collection ref:**	**Capability granted**
Provides for the Comprehensive Representation of Personal Data Processing Activities	Importance of an organisational data catalogue in supporting ERoPA	DO3/ EF5	Recommendation
	ERoPA contained all the terms needed to prepare the mappings, and the cardinality requirements of DPCat ensured that mandatory dataset fields were met.	DO1/ AV5	Positive
	There may be challenges if the organisation model (using the same concepts) takes differing approaches to modelling. Hence, there may be merit in developing best practices for modelling RoPA.	DO5	Recommendation
	The Information held in ROPA alone is not sufficient to demonstrate accountability. There would be significant benefit if other GDPR documents such as data processing agreements and policies and procedures were machine-readable to support accountability verification	AV2 /EF6	Recommendation
Enables Standardised RoPA Information Transfer	ERoPA contained all the terms needed to prepare the mappings, and the cardinality requirements of DPCat ensured that mandatory dataset fields were met.	DO1/ AV5	Positive
	ERoPA interoperability with stakeholders was a strong positive.	EF2 /EF3	Positive
Provides GDPR Reporting Tools for DPO	Tools to support typical DPO tasks in the case study worked successfully	DO2/DO4/AV6/EF1	Positive
Supports Compliance checks & validation	The checking of cardinality rules and GDPR compliance rules using the SHACL specification was successful.	AV3	Positive
Other	An enabling regulator would support the adoption of ERoPA.	EF4	Recommendation
	Stakeholders play an important role in information audits to keep RoPA up-to-date and accurate.	AV1, AV2	Recommendation

7.7 Case Study Findings

The outcome of the case study has identified a number of positives, recommendations, and limitations of ERoPA. This section discusses these findings

Tools and methods: The case study's findings have identified the importance of the CSM-RoPA ontology and DPCat specification specifically developed to support ERoPA in this research. The ontology and specification enable the collection, representation, transfer, and review of ERoPA GDPR information. The use of tools and methods such as (SPARQL, SHACL and OntoRefine) used within the deployment case study all provided integral parts to the deployment of the ERoPA Approach. Each of the tools used played an integral role in the deployment. Alternate tools could be used to conduct these processes; however, the processes supported by these tools must be achieved.

The research has identified the importance of schemas for RDF conversion. Ontorefine was deployed as a tool for Data Conversion to RDF. The conversion to RDF is required to create a schema to support the conversion of data presented as a spreadsheet into RDF format. The Ontorefine tool was used to create a schema and apply the schema to a spreadsheet to convert the data into an RDF format suitable for loading to the knowledge graph.

The importance of a data quality assurance tool for conformance and provenance checking was identified to ensure that input to RoPA provided by stakeholders met data quality rules. The SHACL data shapes ontology was used to create preferred graph shapes, which were used to identify data quality issues and non-compliances. Such conformance and provenance are essential.

The deployment showed a number of standard RoPA tasks, such as a Query Tool, which were required to support the DPO in conducting typical data protection tasks. SPARQL query was used for these processes. SPARQL can export the organisational RoPA to a spreadsheet form or as a DPCat record for exchange with another party system. SPARQL can be used to export data as a regulator template, for inspection by a regulator. It can also be used for ad hoc queries that a DPO may require, such as extracting a particular processing period that is valid at a specific period. The ERoPA Approach has provided a number of standard DPO queries however there may be benefit in the development of a reporting tool or a user interface to support the DPO in this task, to make it usable for normal people.

Data Catalogues: The importance of a Data Catalogue within an organisation became evident in the deployment as the process of collecting and transfer GDPR RoPA information was simplified. The catalogue in the DPCat specification is expressed as the DCAT information system records each processing activity as a dataset, and a RoPA or header information represents a predefined

structure for exchanging information. When data is presented in such a catalogue structure, it is possible to query datasets using specialised tools in a triple store.

Modelling approaches—The Schema created for the case study conversion to RDF was based on how the Privacy Engine software exported the data to a spreadsheet. If the exported data structure changes, the schemas may need to be modified. Also, there may be challenges if the organisation model (using the same concepts) takes differing approaches to modelling. Hence, there may be merit in developing best practices for modelling RoPA. To overcome such challenges, the creation of 'DPV Shapes' that provide suggested data modelling practices for modular use cases. Such shapes, expressed using SHACL, would foster commonality in how the DPV is used and function as a common model for other modelling approaches that can be reduced or aligned. In this, it is important to state that one of the DPV's strengths is its lack of rigid adoption requirements, which provides adopters 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 [18].

ERoPA must be capable of machine-reading other GDPR documents for comparison with RoPA. The expert group agreed that ERoPA is a valuable resource as a stand-alone document. However, to fully benefit from the ERoPA Approach, it must be linked to supporting GDPR documents. This must be tested very carefully. For example, comparing a privacy notice with RoPA, a standard DPO task, may not yet be possible with ERoPA as the language of a RoPA may differ from that of a Privacy Notice, and the same may be true for other GDPR accountability documents. However, a standard vocabulary, such as the DPV, supports this.

An enabling Regulator: The regulator's role in supporting the rollout of ERoPA was identified as beneficial, but it was deemed a lower priority to support its adoption.

7.8 ERoPA Deployment Guidelines

This section builds on the Upsilon deployment case study findings to provide organisations with guidelines for deploying ERoPA. These guidelines are derived from the formalisation of the learning stage of the ADR methodology (see Figure 25) and give organisations the critical guidelines for deploying ERoPA. The goal of section is to gather all the finding that have been gathered in this research, to synthesise this data to establish the key findings, and to present this information in a format to support organisations considering ERoPA deployment. The intended users of this information will be Data protection professionals and technologists; therefore, the information is presented in a format suitable for both groups who will use these guidelines of users.

7.8.1 Methodology

The methodology used to create the guidelines for ERoPA deployment is as follows:

1. Define the objectives and scope of the guidelines
2. Select a suitable framework tailored to present the guidelines
3. Collect data from data sources
4. Analyse the collected data
5. Map the relevant findings into the framework.

Define the objectives and scope of the guidelines: this thesis has gathered findings from many sources and stakeholders as part of the research into ERoPA. Among these sources are as follows:

- State of the Art review (see Section 2.6)
- Survey of Data Protection Professionals (see Section 4.6.2)
- CSM-RoPA implementation (see Section 5.5)
- DPCat-use cases (see Section 6.6)
- ERoPA deployment Case study (See Section 7.7)
    - Observations gathered
    - Expert interviews
    - ICO Accountability Framework implementation

The information collected has been combined into a framework that will provide guidelines to assist organisations considering ERoPA deployment.

Selection of Framework: The framework selected to represent the guidelines is a Zachman Framework (see Section 3.2). The Zachman framework is chosen to support the ERoPA deployment in an organisation. It is an initiative-taking business tool that can model an organisation's existing functions, elements, and processes while helping manage business change. The key benefit of the Zachman Framework is that it provides a holistic perspective on the whole enterprise while allowing focus on specific aspects of the object. For background information on Zachman frameworks, please refer to Section 2.6. The version of the Zachman framework is a Role-Based View of the Zachman Diagram. This type of diagram visualises each row as a perspective or viewpoint specific to different stakeholder roles (e.g., executives, business owners, architects, developers). It helps stakeholders see which parts of the framework are relevant to their roles, clarifying how each viewpoint contributes to the overall enterprise architecture. Focusing on the rows corresponding to specific roles enhances organisational alignment and communication. The role-based diagram serves ERoPA well as it requires cross-functional cooperation and integration to support the collection, representation, and transfer of RoPA information.

7.8.2 Data Collection

Data Collection This section maps the findings from this thesis to the Zachman framework. The findings are presented in Table 55 below, which shows in each finding and which section of the thesis it was identified. Each finding has a unique reference number in the first column, which is used to map to the Zachman framework in Table 56.

Table 55 ERoPA Findings Gathered in this Thesis.

Reference No.	ERoPA Findings gathered from this research.	Where was the finding identified?						
		State of the Art Review	Survey	CSM-RoPA	DPCat Use case	ERoPA deployment observations	Expert interviews	ICO Accountability Verification
F1	Organisations are significantly challenged to maintain accurate and up-to-date RoPA.	✓	✓					
F2	ERoPA must support interoperability with all stakeholders.		✓		✓	✓	✓	
F3	A standard Ontology is required to represent RoPA information.	✓	✓	✓	✓	✓	✓	✓
F4	The DPV vocabulary can fully support the representation of RoPA information concepts.			✓		✓		✓
F5	A triple store such as GraphDB is required to maintain the ERoPA RDF data.			✓	✓	✓		
F6	ERoPA data transfers require a standard specification.		✓		✓	✓	✓	
F7	DPCat fully supports ERoPA data transfers for GDPR RoPA data.				✓	✓		
F8	Importance of organisational Data Catalog to support ERoPA					✓		
F9	ERoPA requires querying tools to support typical DPO tasks.			✓	✓	✓	✓	
F10	SPARQL query can be used to support typical DPO tasks.			✓	✓	✓	✓	
F11	ERoPA requires validation and verification tools to support data quality and provenance.		✓		✓	✓		
F12	SHACL Ontology can support validating and verifying GDPR RoPA information to ensure data quality and provenance.				✓	✓		
F13	ERoPA requires Schemas for RDF conversion.					✓		
F14	OntoRefine can be used as a schema for RDF conversion of GDPR RoPA information.					✓		

Reference No.	ERoPA Findings gathered from this research.	Where was the finding identified?						
		State of the Art Review	Survey	CSM-RoPA	DPCat Use case	ERoPA deployment observations	Expert interviews	ICO Accountability Verification
F15	ERoPA needs to be capable of machine reading other GDPR documents for comparison with RoPA (for example - privacy notice)						✓	✓
F16	An enabling Regulator would support the adoption of ERoPA.	✓	✓				✓	
F17	User Requirements for ERoPA	✓	✓					

7.8.3 Data Analysis:

The section uses the data collection in Table 55 to establish the key findings of the research. There were seventeen findings gathered across the seven sources of (i) State of the Art review (ii) Data Protection Professionals survey (iii) CSM-RoPA development (iv) DPCat use case (v) ERoPA deployment observations (vi) expert interviews and (vii) accountability verification exercise.

The key findings are as follows:

1. While there were seventeen finding in total identified in the research, it was the observations gathered in case study deployment that identified thirteen of these finding. The State-of-the-art review and the Data Protection Professionals survey (see Section 4.6.2) provided seven findings. This shows the benefit of the deployment case study to gather practice-based findings which form a strong basis for future deployments.
2. All seven of the data sources data sources analysed have identified the need for a standard ontology to represent RoPA information. This contrasts very differently from what is happening in practice where vendor software privacy solutions lack open-source ontologies.
3. The DPCat specification supported by the SHACL quality assurance specification provided string a validation and verification process. This is particularly important when data is being exchanged between stakeholders to support data quality and provenance.
4. The case study deployment demonstrated the importance of a query engine to support typical DPO tasks. In the Upsilon deployment a GraphDB triple store with a SPARQL query engine was used. This query tool was critical for the DPO to monitor risk using queries to identify any gaps or conflicts in RoPA. Whilst the case study contained four typical DPOs tasks using SPARQL, for practical use of ERoPA, there would be a requirement to specify

amore queries, or provide a query user interface for the DPO, for comprehensive coverage an ease of use

5. Several of the tools employed in the deployment provided to be very competent tools to support the deployment. Examples of this are the GraphDB triple store and the Ontorefine tool for conversion from spreadsheet to RDF.
6. The Importance of organisational Data Catalog to support ERoPA was identified in the cases study deployment. Whilst this was not identified in any other of the data sources, it was practical observation gathered from the deployment that supported the ease of deployment
7. The expert interviews and the accountability verification sections of the case highlighted the importance that ERoPA must be capable of machine reading heterogeneous GDPR compliance documents for comparison with RoPA (for example - privacy notice). This ability would provide significant benefits for the automated checking of compliance by cross referencing other compliance documents
8. The role of the DPV as an ontology for representing RoPA information concepts was confirmed in the CSM-RoPA development and ERoPA deployment. The DPV also fully supported the concepts required for the ICO accountability framework verification.
9. An enabling regulator would be a catalyst to support the adoption of ERoPA. Whilst this was a critical success factor for RegTech, it is not essential for ERoPA, but the adoption of a regulator supported standardised interoperability would be an enabler.

7.8.4 Approach for Implementing ERoPA Zachman Framework

Mapping into the framework: This section takes the findings gathered through this thesis (see Table 55) and uses these to provide guidelines for organisations deploying ERoPA. These guidelines are gathered from the development rooted in the development of the ERoPA components such as CSM-RoPA and DPCat and from the real-world findings in the Upsilon case study, highlighting lessons learned, challenges encountered and practical solutions. By assessing the deployed artefact, the framework presents refined design guidelines that provide actionable knowledge for organisations considering ERoPA deployment. The findings are communicated to practitioners as practical guidelines and to academics as theoretical insights. The findings are organised and presented in the two-dimensional Role-Based View of a Zachman Diagram, which visualises each row as a perspective or viewpoint specific to planners, business owners, developers, builders, and DPOs (see Table 56). Each entry to the Zachman framework references the relevant from Table 55 using the unique reference in the first column of Table 55.

How the framework should be used: The Zachman Framework (Table 56) is a systematic method for comprehending and recording an organisation's architecture by addressing six essential inquiries (What, How, Where, Who, When, Why) from five distinct viewpoints (Planner, Business Owner, Developer, Builder and DPO). The framework should be used as follows:

1. **Understand the organisational structure:** Each cell in the matrix represents a specific viewpoint and aspect of the enterprise.
2. **Collect comprehensive information:** Fill in each cell to cover all aspects of ERoPA thoroughly.
3. **Analyse for gaps and overlaps:** Identify missing information or redundancies.
4. **Facilitate communication:** Provide a common language for stakeholders to discuss and align on enterprise aspects.
5. **Guide strategic decisions:** Use the information to align strategies with business goals.
6. **Support enterprise architecture development:** Ensure the architecture is integrated and cohesive.
7. **Facilitate change management:** Assess and plan for the impact of changes across the organisation.

This helps in comprehensive documentation, effective communication, and strategic alignment within the enterprise.

Table 56 Guidelines for organisations deploying ERoPA.

	Data/ What	**Function/How**	**Network/ Where**	**People /Who**	**Time/When**	**Motivation/ Why**	
Objective Scope: Planners view	GDPR personal data processing activities (F3, F15)	Collect and represent GDPR accountability data from all potential sources (F2, F3).	GDPR accountability Data stored anywhere is the extended organisation (F2, F15)	All Data Protection Stakeholders, such as Organisation units, Processors, controllers, certification bodies, auditors, the DPO (F2)	As specified by the business owner	To maintain an accurate, comprehensive, and up-to-date ROPA. (F1)	Scope and Context
Requirement Enterprise model: Business Owner View	Interoperability of heterogenous GDPR accountability data (F2)	Common semantic representation and Interoperability of GDPR Information (F3, F4, F6, F7)	Centralised repositories or easily attainable data repositories(F5)	All Data Protection Stakeholders	ROPA must be regularly updated to ensure accurate and comprehensive (F1)	To optimise GDPR compliance by providing an accurate date and comprehensive ROPA, where gaps, conflicts, and non-compliances can be identified (GDPR Art39)	Conceptual
Model of the Information System: (design) Developers view	Logical Data Model, Design Specification, Ontology specification (F3, F4, F6, F7, F8)	Application Architecture, agreed ontology, data sources, and user requirements. (F3, F6, F8, F9, F11,F13)	Distributed system architecture, Data Privacy Vocabulary, Agreed ontology, Organizational data Governance system	Human Interface suitable for users (F9)	Upfront agreement with ontology and specification for transfers (F3, F6)	Business Rule Model Common semantics will enable interoperability and overcome the heterogeneity of data sources. (F3,F4, F6,F7)	Logical

Technology Model. Builders view (plan)	Physical Data Model (F5)	System Design: Schema's, Knowledge Graph, Reporting tools, Data Validation tools (F3, F6, F8, F9, F11, F13)	Technology Architecture: Triple Store (F5)	Presentation architecture suitable for end users such as Data Protection resources and data protection officer (F9)	As specified by the business owner and DPO (F9)	Optimal tools for End users (F9)	Physical
Functioning system: **DPO View**	A RoPA solution that supports typical DPO RoPA tasks to maintain an accurate and up-to-date RoPA (F1, F9)	Collect GDPR accountability data from all potential sources. Represent, validate, and analyse RoPA data (F3, F6, F9)	GDPR Accountability was gathered from all relevant sources and brought to a central repository for analysis (F5, F8)	Organisational Units: Processors, joint controllers, Certification Bodies, Regulators, and the DPO (F2)	ROPA must be up to date; data must be gathered regularly, at agreed intervals, or event-based	To optimise GDPR compliance by providing an accurate date and comprehensive ROPA, where gaps, conflicts, and non-compliances can be identified (GDPR Art39)	Enterprise

7.9 Chapter Conclusion

This chapter addressed two distinct research sub-questions. The first question, RSQ3, was to establish the extent to which the application of the ERoPA supports implementing GDPR accountability.

The Upsilon case study demonstrated the successful deployment of the ERoPA Approach in a real-world organisation. The tools and methods to support the ERoPA case approach were deployed and tested, and four typical DPO tasks were conducted. The DPCat specification for transferring GDPR information between stakeholders enables the successful transfer of GDPR, allowing the stakeholders to interoperate with RoPA. The case study deployment successfully demonstrated that the ERoPA Approach can meet the requirements of GDPR Article 30 obligations regarding RoPA. The RoPA is a critical element in demonstrating GDPR accountability, as it contains a record of the organisation's personal data processing.

Whilst the ERoPA Approach meets the GDPR article 30 RoPA requirements, the challenge for an organisation to meet the Accountability Principle extends beyond the boundary of the RoPA. For the case study deployment, the extent to which ERoPA could meet accountability is limited to the data gathered from the Privacy Engine system. For the ERoPA Approach to fully support the Accountability Principle, other compliance documents, such as privacy notices and data processing agreements, must be maintained in a machine-readable format to be automatically validated against RoPA. This will require organisations to have data governance and catalogue capabilities to support the automation process with metadata. The use of the DPV offers possibilities in this area as it contains many of the GDPR concepts, and the use of DPV Shapes (see Section 7.7) could provide a solution to overcome modelling challenges.

This chapter also addresses RSQ4 to identify the key considerations for organisations implementing an ERoPA Approach. A Zachman framework was provided to support this research sub question based on (i) State of the Art review, (ii) a survey of DPOs, (iii) requirements gathering for ERoPA and the case study findings to support organisations deploying ERoPA.

8 Conclusions

8.1 Chapter Overview

This chapter summarises the research outcomes and reflects on the ADR process for developing the ERoPA Approach artefact. The problem formulation stage of the ADR methodology conceptualised the ERoPA Approach, and it was developed iteratively over three BIE cycles, each involving a design, action, and evaluation cycle. Section 8.2 provides a summary of the key findings, Section 8.3 describes the extent to which this thesis answers the research question and sub-questions, Section 8.4 provides a critical reflection on the ADR process, Section 8.5 provides the contributions of this research, Section 8.6 provides an overview of the limitations of this research and Section 8.7 provides recommendations for future work to develop the ERoPA Approach further.

8.2 Summary of the Key Findings

The ERoPA Approach was developed to help organisations overcome the challenges of maintaining an accurate and up-to-date RoPA. The approach supports organisations with the collection, representation, transfer, and review of RoPA information to demonstrate compliance with the Accountability Principle of the GDPR. The ERoPA Approach contains tools and methods to support DPOs with the collection, representation, transfer, review, and inspection of RoPA information (see Section 4.2). The developed artefact contains (i) a common semantic model of RoPA (CSM-RoPA), which represents all the concepts required to express RoPA information in a machine-readable format, (ii) the DPCat interoperability specification created to collect and transfer RoPA data between stakeholders (iii) SPARQL reporting queries to support typical DPO review tasks, (iv) SHACL shapes to support the validation and conformance of RoPA information, (v) OntoRefine schema for data conversion to RDF (vi) GraphDB as a triple store and (vii) a set of guidelines for considering deployment of the ERoPA Approach.

The key findings indicate that the ERoPA Approach meets the requirements that were established (see Section 4.7) for the collection, representation, transfer, and review of GDPR accountability data for RoPA. The ERoPA Approach was found to provide comprehensive representation of personal data processing activities, enabled standardised RoPA information transfer, supported compliance checks & validation and provided GDPR reporting and review tools for the DPO (See Section 7.6.4). The preparation of the ERoPA Approach deployment guidelines identified the following key findings (see Section 7.8): (i) the importance of a standard ontology to represent RoPA information, contrasting with the current lack of open-source ontologies in vendor software solutions (ii) the DPCat specification, supported by SHACL, ensured a robust

validation process for data exchange, enhancing data quality and provenance (iii) the necessity of a query engine for DPO tasks, using a triple store and SPARQL to monitor risks associated with RoPA, and a user interface are required to facilitate comprehensive usage, (iv) tools like GraphDB and Ontorefine proved effective in supporting deployment activities (v) the Upsilon case study emphasised the benefit that an organisational Data Catalogue made for ERoPA, as this provided a consistent structure for conversion to RDF, and reduced the conversion effort (vi) the expert interviews highlighted the need for GDPR compliance documents (such as privacy notices) to be machine readable, so that they can be compared with the ERoPA information to enhance automated compliance checks (vii) the role of the DPV ontology in representing RoPA concepts was affirmed, and (viii) while an enabling regulator could enhance ERoPA adoption, it is not essential for adoption.

This analysis of the ERoPA Approach identified three ERoPA limitations: (i) the organisation must have a data governance capability to use the ERoPA Approach, [108], [186] (ii) the ERoPA Approach alone does not make an organisation compliant with the Accountability Principle. To realise this, organisations must ensure that all relevant GDPR information is machine-readable to enable interoperability with ERoPA [11], [35], [187] (iii) Another limitation of ERoPA was that organisations might use different modelling approaches because of their unique data requirements and business goals, technology choices, organisational expertise, and legacy integrations [188]The DPV may offer a solution to mitigate this limitation by using DPV shapes to ensure consistency of modelling approaches for RoPA.

8.3 Responding to the Research Question

This section discusses the extent to which this thesis answers the research question and sub-questions addressed. Progress on the research sub-questions is discussed first, followed by progress on the overall research question.

The ADR problem formulation stage successfully addresses **RSQ1:** "*What are the stakeholder requirements for the ERoPA Approach*?" The requirements gathering from authoritative and academic sources and from practice (see Section 4.2) and subsequent analysis (see Section 4.6) identified fourteen requirements for the ERoPA Approach, including the need for an ontology to represent RoPA information, an interoperability specification to enable the exchange of information, and application requirements to enable communication in a machine-readable manner to ensure validity and conformance (see Section 4.7). This was the first time that such an analysis on electronic RoPA requirements has been completed in the state of the art. This was supported by a survey of Data Protection Professionals and validated using the researchers' experience as a practising Data Protection officer, providing a perspective from day-to-day industry practice.

The problem formulation stage also addressed **RSQ2.a,** which" *identifies the information required to be maintained in ERoPA"*. In Chapter 4, an analysis of 17 GDPR regulator RoPA templates published by Data Protection Authorities on their websites was conducted (see Section 4.5.2). This was the first comprehensive analysis of the RoPA templates issued by regulators and it identified that 47 information concepts were required to express a RoPA (see Sections 4.5.2 and 5.2.5). These findings were published in the Legal Knowledge and Information Systems Conference, Jurix [13]. Further refinement of the information gathering process may be possible via native speaker evaluation of non-English language regulator templates.

The first BIE loop concerned the **RSQ2.b** is "*to determine the steps required to develop an ontology for implementing ERoPA information"*. Chapter 5 designs and implements the CSM-RoPA ontology to represent the forty-seven concepts identified from regulator templates. The ERoPA Approach ontology model was created using the NeOn methodology and was developed iteratively over three versions by practitioners to provide a verified representation of all RoPA information concepts found in the regulator RoPA templates. CSM-RoPA version 1 (2020) was presented 33rd International Conference on Legal Knowledge and Information Systems (JURIX 2020) and was also presented at the 23rd International Conference on Enterprise Information Systems, where it was awarded the prize for best paper [33]. CSM-RoPA Version 2 was also published in a journal publication in 2022 [19]. The development of this ontology contributed 26 new terms focused on RoPAs to the Data Privacy Vocabulary (see Section 5.4.4), which are now in use by industry and is the leading vocabulary of data protection and regulation concepts [36].

The second BIE loop expands the ERoPA Approach to build on the CSM-RoPA ontology to provide the DPCat interoperability specification to satisfy **RSQ2c:** "*How can the information required by the ERoPA Approach be communicated between stakeholders?".* The requirements for the ERoPA Approach interoperability specification are identified in the ERoPA requirements specification (see Section 4.7). This was met with the DPCat interoperability specification, which successfully represented five interoperability scenarios between key GDPR stakeholders. The DPCat interoperability specification was presented at the International Conference on Semantic Systems (SEMANTiCS), Amsterdam, Netherlands, in 2021 [34] and was published in the MDPI Journal of Information in 2022 [18].

In the third BIE, a case study deployment of the ERoPA Approach was conducted in a real-world organisation to address **RSQ3:** "*To what extent does the ERoPA Approach support implementing GDPR accountability?"* where a prototype ERoPA Approach was implemented in a real-world environment, and four typical DPO tasks were completed. These standard, typical tasks were identified as part of the Data Protection Professionals survey in Chapter 5.7.2. The case study deployment was presented to data protection experts, who confirmed that the ERoPA Approach is a valuable resource and supports these typical DPO tasks, supporting compliance with the GDPR Accountability principle. The experts suggested that to benefit from the ERoPA Approach fully, it should be linked to supporting GDPR documents. An example of

this, comparing a privacy notice with RoPA, a standard DPO task, may not yet be possible with ERoPA, as the language of a RoPA may differ from that of a Privacy Notice (see Section 7.6.2), and the same may be true for other GDPR accountability documents. However, a standard vocabulary, such as the DPV, supports this, and so by interworking with DPV, the ERoPA Approach supports expansion to these use cases.

The formalisation of the Learning stage of ADR brought together three distinct sources of evidence from the case study as follows: (i) deployment observations, (ii) interviews with five data protection experts who were shown a presentation of the deployment, and (iii) a quantitative study of the completion of the ICO accountability framework using ERoPA. These data sources of evidence were synthesised in Chapter 7 to answer **RSQ4** *"What are the key considerations for organisations implementing an ERoPA Approach?"*. This synthesis is presented in a Zachman framework based on findings gathered from (i) the state-of-the-art review, (ii) a survey of DPOs, (iii) requirements gathering for ERoPA and the case study findings to support organisations deploying ERoPA.

The significant scale of data gathering and analysis to provide key considerations for organising ERoPA deployment ensured that formalisation of the learning stage of ADR was comprehensive in this research. The practitioner's gathering of observations through the case study deployment provided key observations on some of the challenges organisations face with a real-world deployment of ERoPA. Using an expert group to review the ERoPA Approach provided strong feedback and knowledge creation on the case study of ERoPA deployment and opportunities for future work.

The overall research question for this thesis is: '*To what extent can an electronic approach to RoPA support the demonstration of compliance with the Accountability Principle of the GDPR?'.* The ERoPA Approach was iteratively developed through this thesis to meet all requirements established (See Section 4.7). The case study deployment demonstrates the successful completion of typical DPO tasks and the successful population of the ICO Accountability framework based on an ERoPA. Together, it can be seen that these capabilities granted by the ERoPA Approach enable DPOs to demonstrate compliance with the Accountability Principle of the GDPR. However, although the RoPA is a critical GDPR compliance document, adopting the ERoPA Approach alone does not make an organisation compliant, and thus the ERoPA Approach should be seen as support to compliance, hence there is still a need for other aspects of GDPR compliance to be checked when the ERoPA Approach is followed.

8.4 Reflection on the ADR Process

This section reflects on the use of Action Design Research in this study, how the stages of the ADR were applied in this research and how the iterative cycles influenced the evolution of the ERoPA Approach artefact.

Reflections on the Research Process: The objective of this research was to provide an electronic approach to RoPA to support the demonstration of compliance with the GDPR's Accountability Principle. The stages of ADR (Problem Formulation, Building/Intervention/Evaluation, and Formalisation of Learning) cycles were applied to iteratively influence the evolution of the ERoPA Approach artefact.

The problem formulation stage was successful in that the generated comprehensive requirements specification for the ERoPA Approach provided a solid basis for development in the BIE stages. The three data sources, (i) authoritative, (ii) academic, and (iii) practice, used to elicit data requirements provided a solid base for reference through the BIE stages.

The decision to build the ERoPA Approach iteratively, with components developed and added gradually in the BIE stages, worked well. An example of this is that the representation of RoPA was completed in BIE1 first before progressing to more complex steps, like the transfer of RoPA information between stakeholders, which was completed in BIE2. This process complemented the researcher's educational journey, where new tools and methods were identified through these ADR cycles. An example of this is the selection of the DCAT pattern, which was identified to support the interoperability specification, or SHACL, which was identified to support conformance checking of information exchanged between stakeholders.

The decision to conduct a case study on the deployment of ERoPA in an organisational context in BIE3 proved challenging but provided very valuable research findings. The case study deployment proved challenging as there was a lack of Semantic Web skills within the Upsilon organisation. Similarly, the engineering effort required to create a live implementation of the ERoPA Approach was substantial and required significant effort from practitioners and users, with a high dependency on a limited number of people.

The formalisation of learning brought together the findings from the (i) State of the Art review (ii) Data Protection Professionals survey (iii) CSM-RoPA development (iv) DPCat use case (v) ERoPA deployment observations (vi) expert interviews and (vii) accountability verification exercise. This information was successfully presented a **Zachman Framework** [30] (see Section 7.8.3). This stage of ADR worked well as it brought together a significant amount of learning

Stakeholder engagement: The practitioners and user participants for each stage of the ERoPA Approach's development are selected based on their relevance and experience with that stage of the ERoPA Approach.

The selection of users and practitioners is a strength of this research, as they bring a significant amount of practical business experience (see Section 3.1). The blend of legal, technical skills and organisational experience supports the development of the ERoPA Approach through the iterative cycles (Sections 4.1.1, 5.1.1, 6.1.1, and 7.1.1, which set out the specific stakeholder roles for each stage of this research). An example of how the stakeholders provided feedback to optimise the artefact development would be the three mapping cycles for the CSM-RoPA ontology (see Section 5.2.5), where the ontology was iteratively improved in BIE1.

Personal and Researcher Learning: Through this research process, the researcher has gained a significant number of skills. These skills comprise (i) research skills and (ii) technical skills. The research gained skills such as conducting surveys, gaining ethics approval, conducting structured interviews and data analysis skills. The technical skills learned include Semantic Web technology and specific tools such as SHACL, SPARQL and Ontorefine. The researcher did not come from a computer science background and, as such, faced challenges with the technology aspects of this research. However, the benefit of learning these skills enabled the researcher to blend his existing Data Protection knowledge with research and technical skills to research the ERoPA Approach.

Reflections on ADR as a Method: The selection of an ADR method for this research enabled the production of a blended artefact based on theoretical concepts and organisational input. The role of stakeholder input enhanced the artefact's practicality. Through the development of the ERoPA Approach, the structured reflection gathered from users, such as surveys of Data Protection Professionals, the DPVCG, users in the Upsilon organisation and Data Protection Experts, supported improved rigour and insight generation for the development of the ERoPA Approach.

8.5 Contributions

The GDPR mandates that organisations keep a RoPA and ensure compliance. Current RoPA practices rely on spreadsheets or proprietary systems, which lack machine readability and interoperability, creating obstacles to automation. Regulators report that organisations face challenges in maintaining an accurate and up-to-date RoPA. This research analysed the templates provided by GDPR regulators and showed the variation between idealised Art 30 RoPA and the reality of what regulators expect to see recorded in a RoPA. This was the first time this was completed in the state-of-the-art.

The main contribution of this thesis is the ERoPA Approach, which (i) formulates the requirements for a machine-readable RoPA for GDPR based on stakeholder needs (ii) meets these requirements with the CSM-RoPA ontology for RoPA representation and (iii) the DPCat RoPA interoperability specification for exchanging information between stakeholders (iv) provides tools and methods such as RDF conversion, a data catalogue for metadata management, quality assurance control and validation via SHACL (Shapes Constraint Language), and SPARQL query for retrieving required information and (v) deployment guidelines for organisations considering the ERoPA Approach.

The ERoPA Approach enhances GDPR accountability by facilitating the collection, representation, transfer, and review of RoPA information exchanged among stakeholders in data processing chains. The ERoPA Approach enables sharing GDPR accountability information with regulators and certification bodies, significantly improving the visibility and efficiency of organisational accountability practices. Additionally, it provides tools to support GDPR compliance automation. The ERoPA Approach provides data controllers with a comprehensive metadata foundation for creating a machine-readable RoPA that helps Data Protection Officers (DPOs) meet GDPR accountability requirements and identify compliance gaps. DPCat enables stakeholders to share GDPR accountability information efficiently, resulting in a consistent and accurate RoPA across organisations.

The development of the ERoPA Approach improves the Data Privacy Vocabulary (DPV), a state-of-the-art resource for representing GDPR information using semantic web standards. This thesis contributed 26 new terms focused on RoPAs to improve DPV's coverage of GDPR and RoPA information (see Section 5.4.4). The semantic DPCat interoperability specification component of the ERoPA Approach provides an opportunity for organisations to collect, represent, transfer and review GDPR information and surpasses conventional methods that use spreadsheets in terms of information management and potential for automating GDPR compliance, and enables the establishment of communication channels for compliance information amongst stakeholders – a key concept in the EU's vision of Data Spaces [31]. The DPCat approach has now been leveraged to support providers and deployers of high-risk AI applications in registering their

systems in the EU databases. AICat - an extension of DCAT for representing catalogues of AI systems, enables representing catalogues of AI systems that provide consistency, machine-readability, searchability, and interoperability in managing open metadata regarding AI systems [189]. This open approach to cataloguing ensures transparency, traceability, and accountability in AI application markets beyond the immediate needs of high-risk AI compliance in the EU.

8.6 Limitations

The key limitations of this research are (i) the case study was conducted for one organisation only albeit that the Upsilon is a large complex organisation, therefore the results may not be representative of other organisations, and the results might not be transferable to different situations (ii) the ERoPA Approach deployed in the Upsilon Case Study did not provide Graphic User Interfaces (GUIs) for end users. Organisations wishing to deploy ERoPA in a commercial context would require such user interfaces, which would require development (iii) the Upsilon Case Study did not use commercial-grade engineering in its deployment. For future commercial deployments, organisations would need to adhere to industry best practices and international standards and focus on designing, developing, and delivering a robust, scalable, and maintainable ERoPA solution (iv) the ERoPA Approach does not involve any specific new technology in live use. In the Upsilon case study, the ERoPA Approach was implemented in a test environment, and (v) the expert group consisted of five experts with significant experience, but the size of the sample group may be considered to be small and could be extended in future work.

8.7 Future Work

The ERoPA Approach effectively enables organisations to create and maintain a machine-readable Record of Processing Activities (RoPA) to support demonstrating GDPR accountability. This lays a strong foundation for other GDPR RegTech tools that can identify GDPR risks and facilitate regular compliance checks. This would require the semantic modelling of other GDPR documents, such as data processing agreements and privacy notices, which extend beyond the RoPA domain. This would require the organisation to develop a data governance model to support this process. The continuous monitoring of such GDPR accountability information could help inform the Data Protection Officer (DPO) of new or modified data processing activities to monitor compliance, supporting the change management process.

ERoPA's capacity to extend beyond the organisation's boundary can facilitate the digital exchange of compliance information between stakeholders, such as processors and controllers, thus building trust and confidence in the data processing chain. The ability to electronically exchange GDPR accountability information with regulators and certification bodies has the added benefit of supporting accountability when

accompanied by some form of external validation, which ensures both demonstration and verification. The capability of the machine-readable RoPA to support the sharing of accountability information with such certification bodies would be a significant benefit to improving the visibility of the organisation's accountability practices. ERoPA exceeds traditional spreadsheet approaches for information management and enhances the ability to automate GDPR compliance. It also facilitates the creation of communication channels for compliance-related information among stakeholders, aligning with the EU's vision for Data Spaces [31].

References

[1] Sein, Henfridsson, Purao, Rossi, and Lindgren, 'Action Design Research', *MIS Quarterly*, vol. 35, no. 1, p. 37, 2011, doi: 10.2307/23043488.

[2] L. R. Baker, 'The ontology of artifacts', *Philosophical Explorations*, vol. 7, no. 2, pp. 99–111, Jun. 2004, doi: 10.1080/13869790410001694462.

[3] G. G. Inan and U. S. Bititci, 'Understanding Organizational Capabilities and Dynamic Capabilities in the Context of Micro Enterprises: A Research Agenda', *Procedia - Social and Behavioral Sciences*, vol. 210, pp. 310–319, Dec. 2015, doi: 10.1016/j.sbspro.2015.11.371.

[4] N. O'Regan and A. Ghobadian, 'The importance of capabilities for strategic direction and performance', *Management Decision*, vol. 42, no. 2, pp. 292–313, Jan. 2004, doi: 10.1108/00251740410518525.

[5] T. Berners-Lee, J. Hendler, and O. Lassila, 'The Semantic Web: A New Form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities', in *Linking the World's Information*, 1st ed., O. Seneviratne and J. Hendler, Eds., New York, NY, USA: ACM, 2023, pp. 91–103. doi: 10.1145/3591366.3591376.

[6] H. J. Pandit, B. Esteves, G. P. Krog, P. Ryan, D. Golpayegani, and J. Flake, 'Data Privacy Vocabulary (DPV) – Version 2.0', in *The Semantic Web – ISWC 2024*, G. Demartini, K. Hose, M. Acosta, M. Palmonari, G. Cheng, H. Skaf-Molli, N. Ferranti, D. Hernández, and A. Hogan, Eds., Cham: Springer Nature Switzerland, 2025, pp. 171–193. doi: 10.1007/978-3-031-77847-6_10.

[7] 'W3C Semantic Web FAQ'. Accessed: Apr. 10, 2024. [Online]. Available: https://www.w3.org/2001/sw/SW-FAQ#WhatIsTheSW

[8] N. Shadbolt, T. Berners-Lee, and W. Hall, 'The Semantic Web Revisited', *IEEE Intell. Syst.*, vol. 21, no. 3, pp. 96–101, May 2006, doi: 10.1109/MIS.2006.62.

[9] D. Torre, M. Alferez, G. Soltana, M. Sabetzadeh, and L. Briand, 'Model Driven Engineering for Data Protection and Privacy: Application and Experience with GDPR', Jul. 23, 2020, *arXiv*: arXiv:2007.12046. doi: 10.48550/arXiv.2007.12046.

[10] D. Korff and M. Georges, 'The Data Protection Officer Handbook', Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3428957, Jul. 2019. Accessed: Mar. 10, 2022. [Online]. Available: https://papers.ssrn.com/abstract=3428957

[11] 'New DPC Guidance on Records of Processing Activities', Default. Accessed: Jan. 07, 2024. [Online]. Available: https://www.matheson.com/insights/detail/new-dpc-guidance-on-records-of-processing-activities

[12] 'Measuring Privacy Operations 2019 Cookies, Local vs. Global Compliance, DSARs and more - IAPP and TrustArc (Accessed: 2022-03-10)'. Accessed: Mar. 10, 2022. [Online]. Available: https://iapp.org/media/pdf/resource_center/measuring_privacy_operations_2019.pdf

[13] P. Ryan, H. J. Pandit, and R. Brennan, 'A Common Semantic Model of the GDPR Register of Processing Activities', *arXiv:2102.00980 [cs]*, Dec. 2020, doi: 10.3233/FAIA200876.

[14] O. Amaral, M. I. Azeem, S. Abualhaija, and L. C. Briand, 'NLP-based Automated Compliance Checking of Data Processing Agreements against GDPR', Sep. 20, 2022, *arXiv*: arXiv:2209.09722. Accessed: Sep. 26, 2022. [Online]. Available: http://arxiv.org/abs/2209.09722

[15] T. Butler and L. O'Brien, 'Understanding RegTech for Digital Regulatory Compliance', in *Disrupting Finance*, T. Lynn, J. G. Mooney, P. Rosati, and M. Cummins, Eds., Cham: Springer International Publishing, 2019, pp. 85–102. doi: 10.1007/978-3-030-02330-0_6.

[16] R. P. Buckley, D. W. Arner, D. A. Zetzsche, and R. H. Weber, 'The road to RegTech: the (astonishing) example of the European Union', *J Bank Regul*, vol. 21, no. 1, pp. 26–36, Mar. 2020, doi: 10.1057/s41261-019-00104-1.

[17] P. Ryan, M. Crane, and R. Brennan, 'GDPR Compliance Tools: Best Practice from RegTech', in *Enterprise Information Systems*, J. Filipe, M. Śmiałek, A. Brodsky, and S. Hammoudi, Eds., in

Lecture Notes in Business Information Processing. Cham: Springer International Publishing, 2021, pp. 905–929. doi: 10.1007/978-3-030-75418-1_41.
[18] P. Ryan, R. Brennan, and H. J. Pandit, 'DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR', *Information*, vol. 13, no. 5, Art. no. 5, May 2022, doi: 10.3390/info13050244.
[19] P. Ryan and R. Brennan, 'Support for enhanced GDPR accountability with the common semantic model for ROPA (CSM-ROPA)', *SN Computer Science*, vol. 3, Apr. 2022, doi: 10.1007/s42979-022-01099-9.
[20] M. M. Martínez-González, P. Casanovas, M.-L. Alvite-Díez, N. Casellas, A. Aparicio, and D. Sanz, 'Privacy Compliance with Ontologies and Blockchain: The OntoROPA Project'.
[21] D. Huth, A. Tanakol, and F. Matthes, 'Using Enterprise Architecture Models for Creating the Record of Processing Activities (Art. 30 GDPR)', in *2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC)*, Oct. 2019, pp. 98–104. doi: 10.1109/EDOC.2019.00021.
[22] 'Principles of Data Protection | Data Protection Commission', Principles of Data Protection | Data Protection Commission. Accessed: Mar. 26, 2024. [Online]. Available: https://www.dataprotection.ie/individuals/data-protection-basics/principles-data-protection
[23] M. C. Suárez-Figueroa, 'NeOn Methodology for Building Ontology Networks:Specification, Scheduling and Reuse', PhD Thesis, Universidad Politécnica de Madrid, 2010. doi: 10.20868/UPM.thesis.3879.
[24] F. Scharffe, 'Correspondence patterns representation', Mar. 2009, Accessed: Mar. 10, 2022. [Online]. Available: https://www.academia.edu/2867981/Correspondence_patterns_representation
[25] 'OWL - Semantic Web Standards'. Accessed: Feb. 04, 2025. [Online]. Available: https://www.w3.org/OWL/
[26] 'Horridge - A Practical Guide To Building OWL Ontologies Using.pdf'. Accessed: Aug. 02, 2022. [Online]. Available: http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_2.pdf
[27] D. Garijo, 'WIDOCO: A Wizard for Documenting Ontologies', in *The Semantic Web – ISWC 2017*, vol. 10588, C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, and J. Heflin, Eds., in Lecture Notes in Computer Science, vol. 10588. , Cham: Springer International Publishing, 2017, pp. 94–102. doi: 10.1007/978-3-319-68204-4_9.
[28] M. D. Wilkinson *et al.*, 'The FAIR Guiding Principles for scientific data management and stewardship', *Sci Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.
[29] S. J. Morgan, S. R. H. Pullon, L. M. Macdonald, E. M. McKinlay, and B. V. Gray, 'Case Study Observational Research: A Framework for Conducting Case Study Research Where Observation Data Are the Focus', *Qual Health Res*, vol. 27, no. 7, pp. 1060–1068, Jun. 2017, doi: 10.1177/1049732316649160.
[30] J. A. Zachman, 'The zachman framework for enterprise architecture', *Primer for Enterprise Engineering and Manufacturing.[si]: Zachman International*, 2003.
[31] 'What are data spaces? Systematic survey and future outlook - ScienceDirect'. Accessed: Feb. 22, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340924009314?via%3Dihub
[32] P. Ryan, M. Crane, and R. Brennan, 'Design Challenges for GDPR RegTech':, in *Proceedings of the 22nd International Conference on Enterprise Information Systems*, Prague, Czech Republic: SCITEPRESS - Science and Technology Publications, 2020, pp. 787–795. doi: 10.5220/0009464507870795.
[33] P. Ryan and R. Brennan, 'Demonstrating GDPR accountability with CSM-ROPA: extensions to the data privacy vocabulary', in *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.*, Online: ICEIS, Apr. 2021, p.

Accessed: Mar. 10, 2022. [Online]. Available: https://www.insticc.org/node/TechnicalProgram/iceis/2021/presentationDetails/103905
[34] H. Pandit, 'Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV', 2021. Accessed: Mar. 10, 2022. [Online]. Available: http://www.tara.tcd.ie/handle/2262/96594
[35] B. Gonçalves Crisóstomo Esteves, H. J. Pandit, G. P. Krog, and P. Ryan, 'How to manage my data? With machine–interpretable GDPR rights!', in *JURIX 2024 : the Thirty-seventh Annual Conference, Brno*, IOS Press, 2024, pp. 269–274. doi: 10.3233/faia241254.
[36] 'Signatu'. Accessed: Jan. 17, 2025. [Online]. Available: https://signatu.com/home/
[37] 'ISO/IEC PWI TS 27569'. Accessed: Jan. 12, 2025. [Online]. Available: https://genorma.com/en/standards/iso-iec-pwi-ts-27569
[38] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, 'The NeOn Methodology for Ontology Engineering', in *Ontology Engineering in a Networked World*, M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–34. doi: 10.1007/978-3-642-24794-1_2.
[39] M. Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework1', *European Law Journal*, vol. 13, no. 4, pp. 447–468, 2007, doi: 10.1111/j.1468-0386.2007.00378.x.
[40] 'Opinion 3/2010 on the principle of accountability (EDPB)'. Accessed: Nov. 28, 2024. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp173_en.pdf
[41] C. Labadie and C. Legner, 'Understanding Data Protection Regulations from a Data Management Perspective: A Capability-Based Approach to EU-GDPR', presented at the Wirtschaftsinformatik, 2019, p. 15. [Online]. Available: https://aisel.aisnet.org/wi2019/track11/papers/3/
[42] 'cipl_accountability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf'. Accessed: Mar. 30, 2024. [Online]. Available: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_paper_1_-_the_case_for_accountability_-_how_it_enables_effective_data_protection_and_trust_in_the_digital_society.pdf
[43] 'Accountability obligation | Data Protection Commission', Accountability obligation | Data Protection Commission. Accessed: Mar. 30, 2024. [Online]. Available: https://www.dataprotection.ie/organisations/know-your-obligations/accountability-obligation
[44] E. Lachaud, 'Accountability and Certification in the GDPR', Oct. 22, 2021, *Social Science Research Network, Rochester, NY*: 3948093. doi: 10.2139/ssrn.3948093.
[45] 'Privacy seals, certifications & marks as a result of the GDPR | Cyber Security | Privacy', Deloitte Netherlands. Accessed: Apr. 08, 2024. [Online]. Available: https://www2.deloitte.com/nl/nl/pages/risk/articles/cyber-security-gdpr-privacy-seals-certifications-and-marks.html
[46] 'Accountability'. Accessed: Feb. 15, 2025. [Online]. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/data-sharing-a-code-of-practice/accountability/
[47] 'The value of investing in well-constructed records of processing activities - IAPP (Accessed: 2022-03-10)'. Accessed: Mar. 10, 2022. [Online]. Available: https://iapp.org/news/a/the-value-of-investing-in-well-constructed-recordings-of-processing-activities/
[48] 'Record of processing activities | CNIL'. Accessed: Mar. 10, 2022. [Online]. Available: https://www.cnil.fr/en/record-processing-activities
[49] 'Records of processing and lawful basis - ICO (Accessed: 2022-03-10)'. Accessed: Mar. 10, 2022. [Online]. Available: https://ico.org.uk/for-organisations/accountability-framework/records-of-processing-and-lawful-basis/
[50] 'Guidelines 07/2020 on the concepts of controller and processor in the GDPR | European Data Protection Board'. Accessed: Dec. 17, 2024. [Online]. Available:

https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-072020-concepts-controller-and_en
[51] D. Drewer and V. Miladinova, 'The canary in the data mine', *Computer Law & Security Review*, vol. 34, no. 4, pp. 806–815, Aug. 2018, doi: 10.1016/j.clsr.2018.05.019.
[52] 'Data protection officers'. Accessed: Dec. 17, 2024. [Online]. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/data-protection-officers/
[53] 'Top 10 operational impacts of the GDPR: Part 2 - The mandatory DPO | IAPP'. Accessed: Feb. 15, 2025. [Online]. Available: https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-2-the-mandatory-dpo
[54] 'Record of processing activities — Are you ready for maturity?' Accessed: Aug. 30, 2022. [Online]. Available: https://iapp.org/news/a/record-of-processing-activities-are-you-ready-for-maturity/
[55] 'Demonstrate your compliance with data protection regulations | Data Protection Ombudsman's Office', Tietosuojavaltuutetun toimisto. Accessed: Feb. 07, 2025. [Online]. Available: https://tietosuoja.fi/en/accountability?utm_source=chatgpt.com
[56] 'dpia-template-v04-post-comms-review-20180308.pdf'. Accessed: Apr. 08, 2024. [Online]. Available: https://ico.org.uk/media/2258461/dpia-template-v04-post-comms-review-20180308.pdf
[57] 'Self-Assessment Checklist | Data Protection Commission', Self-Assessment Checklist | Data Protection Commission. Accessed: Apr. 08, 2024. [Online]. Available: https://www.dataprotection.ie/organisations/resources-organisations/self-assessment-checklist
[58] DPM, 'Records of Processing Activities [Templates and Examples for different Industries]', Data Privacy Manager. Accessed: Feb. 07, 2025. [Online]. Available: https://dataprivacymanager.net/how-to-document-records-of-processing-activities-templates-and-examples-for-different-industries/
[59] M. Vartabedian, 'AI Can Take the Slog Out of Compliance Work, but Executives Not Ready to Fully Trust It', *Wall Street Journal*, Dec. 17, 2024. Accessed: Feb. 02, 2025. [Online]. Available: https://www.wsj.com/articles/ai-can-take-the-slog-out-of-compliance-work-but-executives-not-ready-to-fully-trust-it-7cd60a16
[60] '2020TechVendorReport.pdf'. Accessed: Mar. 10, 2022. [Online]. Available: https://iapp.org/media/pdf/resource_center/2020TechVendorReport.pdf
[61] 'Data Privacy Software Market Size, Share & Growth [2032]'. Accessed: Dec. 31, 2024. [Online]. Available: https://www.fortunebusinessinsights.com/data-privacy-software-market-105420
[62] '2022TechVendorReport.pdf'. Accessed: Oct. 13, 2022. [Online]. Available: https://iapp.org/media/pdf/resource_center/2022TechVendorReport.pdf
[63] '2021TechVendorReport.pdf'. Accessed: Mar. 10, 2022. [Online]. Available: https://iapp.org/media/pdf/resource_center/2021TechVendorReport.pdf
[64] 'BestPractices_DataManagement_v1.0.pdf'. Accessed: Jan. 20, 2025. [Online]. Available: https://www.eurocc-access.eu/wp-content/uploads/2022/09/BestPractices_DataManagement_v1.0.pdf
[65] 'aicpa_cica_privacy_maturity_model.pdf'. Accessed: Apr. 08, 2024. [Online]. Available: https://vvena.nl/wp-content/uploads/2018/04/aicpa_cica_privacy_maturity_model.pdf
[66] 'La CNIL propose une autoévaluation de maturité en gestion de la protection des données'. Accessed: Apr. 08, 2024. [Online]. Available: https://www.cnil.fr/fr/la-cnil-propose-une-autoevaluation-de-maturite-en-gestion-de-la-protection-des-donnees
[67] Z. Spalevic and K. Vićentijević, 'GDPR and challenges of personal data protection', *The European Journal of Applied Economics*, vol. 19, pp. 55–65, Jan. 2022, doi: 10.5937/EJAE19-36596.
[68] 'UK ICO Publishes Its Accountability Framework'. Accessed: Feb. 07, 2025. [Online]. Available: https://natlawreview.com/article/ico-publishes-its-accountability-framework

[69] ‘GDPR Certification | Data Protection Commission’, GDPR Certification | Data Protection Commission. Accessed: Feb. 02, 2025. [Online]. Available: https://www.dataprotection.ie/organisations/gdpr-certification
[70] ‘edpb_guidelines_201801_v3.0_certificationcriteria_annex2_en.pdf’. Accessed: Feb. 02, 2025. [Online]. Available: https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201801_v3.0_certificationcriteria_annex2_en.pdf
[71] ‘4 GDPR-certification myths dispelled’. Accessed: Apr. 08, 2024. [Online]. Available: https://iapp.org/news/a/four-gdpr-certification-myths-dispelled/
[72] ‘Register of certification mechanisms, seals and marks | European Data Protection Board’. Accessed: Feb. 07, 2025. [Online]. Available: https://www.edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en
[73] V. Colaert and K. Leuven, ‘RegTech as a response to regulatory expansion in the financial sector’, 2017.
[74] ‘data-protection-officer-survey-2020-21.pdf’. Accessed: Jun. 01, 2022. [Online]. Available: https://www.datatilsynet.no/contentassets/b5f70248207a4768a3295aaffac78edc/data-protection-officer-survey-2020-21.pdf
[75] D. W. Arner, J. N. Barberis, and R. P. Buckley, ‘The emergence of RegTech 2.0: From know your customer to know your data’, 2016, Accessed: Nov. 28, 2024. [Online]. Available: https://papers.ssrn.com/SOL3/PAPERS.CFM?ABSTRACT_ID=3044280
[76] ‘Anti-Money Laundering and Countering the Financing of Terrorism | Central Bank of Ireland’. Accessed: Jan. 20, 2025. [Online]. Available: https://www.centralbank.ie/regulation/anti-money-laundering-and-countering-the-financing-of-terrorism
[77] ‘MiFID II | European Securities and Markets Authority’. Accessed: Jan. 20, 2025. [Online]. Available: https://www.esma.europa.eu/publications-and-data/interactive-single-rulebook/mifid-ii
[78] ‘in-risk-RegTech-Gaining-momentum-noexp.pdf’. Accessed: Feb. 05, 2025. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-risk-RegTech-Gaining-momentum-noexp.pdf
[79] D. Basu and G. K. Tetteh, ‘Using Automation and AI to Combat Money Laundering’, University of Strathclyde, 2024. doi: 10.17868/STRATH.00089571.
[80] P. Kavassalis, H. Stieber, H. Breymann, K. Saxton, and F. Gross, ‘An innovative RegTech approach to financial risk monitoring and supervisory reporting’, *The Journal of Risk Finance*, vol. 19, pp. 00–00, Nov. 2017, doi: 10.1108/JRF-07-2017-0111.
[81] V. Jayagopal and B. K. K, ‘Data Management and Big Data Analytics: Data Management in Digital Economy’, in *Research Anthology on Big Data Analytics, Architectures, and Applications*, IGI Global, 2022, pp. 1614–1633. doi: 10.4018/978-1-6684-3662-2.ch078.
[82] E. Ponick and G. Wieczorek, ‘Artificial Intelligence in Governance, Risk and Compliance: Results of a study on potentials for the application of artificial intelligence (AI) in governance, risk and compliance (GRC)’, May 08, 2024, *arXiv*: arXiv:2212.03601. doi: 10.48550/arXiv.2212.03601.
[83] D. Arner, J. Barberis, and R. Buckley, ‘FinTech, regTech, and the reconceptualization of financial regulation’, *Northwestern Journal of International Law and Business*, vol. 37, pp. 373–415, Jan. 2017.
[84] I. Anagnostopoulos, ‘Fintech and regtech: Impact on regulators and banks’, *Journal of Economics and Business*, vol. 100, pp. 7–25, Nov. 2018, doi: 10.1016/j.jeconbus.2018.07.003.
[85] D. W. Arner, J. Barberis, and R. P. Buckley, *FinTech and RegTech in a Nutshell, and the Future in a Sandbox*. CFA Institute Research Foundation, 2017.
[86] L. G. Baxter, ‘Adaptive Financial Regulation and RegTech: A Concept Article on Realistic Protection for Victims of Bank Failures’, *DUKE LAW JOURNAL*, vol. 66.
[87] N. G. Packin, ‘RegTech, compliance and technology judgment rule’, *Chi.-Kent L. Rev.*, vol. 93, p. 193, 2018.
[88] K. Larsen and S. Gilani, ‘RegTech is the New Black - The Growth of RegTech Demand and Investment’, *Journal of Financial Transformation*, vol. 45, pp. 22–29, 2017.

[89] A.-J. Aberkane, G. Poels, and S. V. Broucke, 'Exploring Automated GDPR-Compliance in Requirements Engineering: A Systematic Mapping Study', *IEEE Access*, vol. 9, pp. 66542–66559, 2021, doi: 10.1109/ACCESS.2021.3076921.
[90] Z. S. Li, C. Werner, N. Ernst, and D. Damian, 'GDPR Compliance in the Context of Continuous Integration', Feb. 17, 2020, *arXiv*: arXiv:2002.06830. Accessed: Apr. 26, 2024. [Online]. Available: http://arxiv.org/abs/2002.06830
[91] D. R. Amariles, A. C. Troussel, and R. E. Hamdani, 'Compliance Generation for Privacy Documents under GDPR: A Roadmap for Implementing Automation and Machine Learning', Dec. 23, 2020, *arXiv*: arXiv:2012.12718. Accessed: Apr. 26, 2024. [Online]. Available: http://arxiv.org/abs/2012.12718
[92] A. Tsohou *et al.*, 'Privacy, security, legal and technology acceptance elicited and consolidated requirements for a GDPR compliance platform', *Information & Computer Security*, vol. 28, no. 4, pp. 531–553, Jan. 2020, doi: 10.1108/ICS-01-2020-0002.
[93] 'Artificial Intelligence-enabled Automation for Compliance Checking against GDPR.pdf'. Accessed: Apr. 26, 2024. [Online]. Available: https://orbilu.uni.lu/bitstream/10993/56026/1/Artificial%20Intelligence-enabled%20Automation%20for%20Compliance%20Checking%20against%20GDPR.pdf
[94] D. S. Guamán, D. Rodriguez, J. M. del Alamo, and J. Such, 'Automated GDPR compliance assessment for cross-border personal data transfers in android applications', *Computers & Security*, vol. 130, p. 103262, Jul. 2023, doi: 10.1016/j.cose.2023.103262.
[95] R. el Hamdani, M. Mustapha, D. Restrepo, A. Troussel, S. Meeùs, and K. Krasnashchok, *A combined rule-based and machine learning approach for automated GDPR compliance checking*. 2021, p. 49. doi: 10.1145/3462757.3466081.
[96] M. Barati, O. Rana, I. Petri, and G. Theodorakopoulos, 'GDPR Compliance Verification in Internet of Things', *IEEE Access*, vol. 8, pp. 119697–119709, 2020, doi: 10.1109/ACCESS.2020.3005509.
[97] T. R. Chhetri, A. Kurteva, R. J. DeLong, R. Hilscher, K. Korte, and A. Fensel, 'Data Protection by Design Tool for Automated GDPR Compliance Verification Based on Semantically Modeled Informed Consent', *Sensors (Basel)*, vol. 22, no. 7, p. 2763, Apr. 2022, doi: 10.3390/s22072763.
[98] T. Libal, 'Towards Automated GDPR Compliance Checking', in *Trustworthy AI - Integrating Learning, Optimization and Reasoning*, F. Heintz, M. Milano, and B. O'Sullivan, Eds., Cham: Springer International Publishing, 2021, pp. 3–19. doi: 10.1007/978-3-030-73959-1_1.
[99] D. Torre, G. Soltana, M. Sabetzadeh, L. C. Briand, Y. Auffinger, and P. Goes, 'Using Models to Enable Compliance Checking Against the GDPR: An Experience Report', in *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS)*, Munich, Germany: IEEE, Sep. 2019, pp. 1–11. doi: 10.1109/MODELS.2019.00-20.
[100] J. Kingston, 'Using artificial intelligence to support compliance with the general data protection regulation', *Artif Intell Law*, vol. 25, no. 4, pp. 429–443, Dec. 2017, doi: 10.1007/s10506-017-9206-9.
[101] A. Mahindrakar and K. P. Joshi, 'Automating GDPR Compliance using Policy Integrated Blockchain', in *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, Baltimore, MD, USA: IEEE, May 2020, pp. 86–93. doi: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00026.
[102] M. Barati, G. Theodorakopoulos, and O. Rana, 'Automating GDPR Compliance Verification for Cloud-hosted Services', in *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, Montreal, QC, Canada: IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/ISNCC49221.2020.9297309.
[103] E. Arfelt, D. Basin, and S. Debois, 'Monitoring the GDPR', in *Computer Security – ESORICS 2019*, vol. 11735, K. Sako, S. Schneider, and P. Y. A. Ryan, Eds., in Lecture Notes in Computer Science, vol. 11735. , Cham: Springer International Publishing, 2019, pp. 681–699. doi: 10.1007/978-3-030-29959-0_33.

[104] R. Zaman and M. Hassani, 'On Enabling GDPR Compliance in Business Processes Through Data-Driven Solutions', *SN COMPUT. SCI.*, vol. 1, no. 4, p. 210, Jun. 2020, doi: 10.1007/s42979-020-00215-x.
[105] M. Brodin, 'A Framework for GDPR Compliance for Small- and Medium-Sized Enterprises', *Eur J Secur Res*, vol. 4, no. 2, pp. 243–264, Oct. 2019, doi: 10.1007/s41125-019-00042-z.
[106] L. Elluri, A. Nagar, and K. P. Joshi, 'An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance', in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA: IEEE, Dec. 2018, pp. 1266–1271. doi: 10.1109/BigData.2018.8622236.
[107] R. Matulevičius, J. Tom, K. Kala, and E. Sing, 'A Method for Managing GDPR Compliance in Business Processes', in *Advanced Information Systems Engineering*, vol. 386, N. Herbaut and M. La Rosa, Eds., in Lecture Notes in Business Information Processing, vol. 386. , Cham: Springer International Publishing, 2020, pp. 100–112. doi: 10.1007/978-3-030-58135-0_9.
[108] V. Khatri and C. V. Brown, 'Designing data governance', *Commun. ACM*, vol. 53, no. 1, pp. 148–152, Jan. 2010, doi: 10.1145/1629175.1629210.
[109] 'NIFO - National Interoperability Framework Observatory | Joinup'. Accessed: May 04, 2024. [Online]. Available: https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory
[110] M. Dumontier *et al.*, 'Consent through the lens of semantics: State of the art survey and best practices', *Semantic Web*, vol. 15, no. 3, pp. 647–673, May 2024, doi: 10.3233/SW-210438.
[111] H. J. Pandit, 'Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance', p. 238.
[112] J. Anim, L. Robaldo, and A. Z. Wyner, 'A SHACL-Based Approach for Enhancing Automated Compliance Checking with RDF Data', *Information*, vol. 15, no. 12, Art. no. 12, Dec. 2024, doi: 10.3390/info15120759.
[113] B. Esteves and V. Rodríguez-Doncel, 'Analysis of ontologies and policy languages to represent information flows in GDPR', *Semantic Web*, vol. Preprint, no. Preprint, pp. 1–35, Jan. 2022, doi: 10.3233/SW-223009.
[114] 'Semantic Interoperability Community', GitHub. Accessed: Jan. 17, 2025. [Online]. Available: https://github.com/SEMICeu
[115] M. M. Martínez González, M. L. Alvite Díez, P. Casanovas, N. Casellas, D. Sanz, and A. Aparicio de la Fuente, 'OntoROPA Deliverable 1. State of the Art and Ambition.', 2021, Accessed: Mar. 10, 2022. [Online]. Available: https://uvadoc.uva.es/handle/10324/47863
[116] J. E. Blumenstock and N. Kohli, 'Big Data Privacy in Emerging Market Fintech and Financial Services: A Research Agenda', 2023, doi: 10.26085/C3WK53.
[117] K. Lyytinen, B. Weber, M. C. Becker, and B. T. Pentland, 'Digital twins of organization: implications for organization design', *J Org Design*, vol. 13, no. 3, pp. 77–93, Sep. 2024, doi: 10.1007/s41469-023-00151-z.
[118] 'Wiki for Privacy Standards and Privacy Projects - IPEN Wiki'. Accessed: Jan. 17, 2025. [Online]. Available: https://ipen.trialog.com/wiki/Wiki_for_Privacy_Standards_and_Privacy_Projects
[119] 'EDPB launches website auditing tool | European Data Protection Board'. Accessed: Jan. 17, 2025. [Online]. Available: https://www.edpb.europa.eu/news/news/2024/edpb-launches-website-auditing-tool_en
[120] P. Rozehnal and V. Novák, 'The Core Of Enterprise Architecture As A Management Tool: Gdpr Implementation Case Study', presented at the 26th Interdisciplinary Information Management Talks, Oct. 2020. [Online]. Available: http://idimt.org/wp-content/uploads/proceedings/IDIMT_proceedings_2018.pdf
[121] F. Burmeister, P. Drews, and I. Schirmer, 'A Privacy-driven Enterprise Architecture Meta-Model for Supporting Compliance with the General Data Protection Regulation', presented at the Hawaii International Conference on System Sciences 2019 (HICSS-52), Jan. 2019. [Online]. Available: http://hdl.handle.net/10125/60040
[122] 'Olympe'. Accessed: Jan. 17, 2025. [Online]. Available: https://www.olympe.legal/

[123] 'UROPA/docs_rst/introduction.rst at master · loosolab/UROPA', GitHub. Accessed: Jan. 17, 2025. [Online]. Available: https://github.com/loosolab/UROPA/blob/master/docs_rst/introduction.rst
[124] 'OpenAPI Specification v3.1.0'. Accessed: Jan. 17, 2025. [Online]. Available: https://spec.openapis.org/oas/v3.1.0.html
[125] B. Esteves and V. Rodriguez-Doncel, 'Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR', *Semantic Web J.*, vol. Forthcoming, 2022, Accessed: Mar. 10, 2022. [Online]. Available: http://www.semantic-web-journal.net/content/analysis-ontologies-and-policy-languages-represent-information-flows-gdpr-1
[126] 'GDPRov - The GDPR Provenance Ontology'. Accessed: Jan. 09, 2025. [Online]. Available: https://openscience.adaptcentre.ie/ontologies/GDPRov/docs/ontology
[127] H. J. Pandit, C. Debruyne, D. O'Sullivan, and D. Lewis, 'GConsent - A Consent Ontology Based on the GDPR', in *The Semantic Web*, P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. Lopez, A. Haller, and K. Hammar, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 270–282. doi: 10.1007/978-3-030-21348-0_18.
[128] 'Home - SPECIAL'. Accessed: Nov. 28, 2024. [Online]. Available: https://specialprivacy.ercim.eu/
[129] H. J. Pandit, K. Fatema, D. O'Sullivan, and D. Lewis, 'GDPRtEXT - GDPR as a Linked Data Resource', in *The Semantic Web*, vol. 10843, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., in Lecture Notes in Computer Science, vol. 10843. , Cham: Springer International Publishing, 2018, pp. 481–495. doi: 10.1007/978-3-319-93417-4_31.
[130] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, and L. Robaldo, 'PrOnto: Privacy Ontology for Legal Reasoning', in *Electronic Government and the Information Systems Perspective*, vol. 11032, A. Kő and E. Francesconi, Eds., in Lecture Notes in Computer Science, vol. 11032. , Cham: Springer International Publishing, 2018, pp. 139–152. doi: 10.1007/978-3-319-98349-3_11.
[131] G. V. Lioudakis *et al.*, 'Facilitating GDPR Compliance: The H2020 BPR4GDPR Approach', in *Digital Transformation for a Sustainable Society in the 21st Century*, I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie, and M. Mäntymäki, Eds., in IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, 2020, pp. 72–78. doi: 10.1007/978-3-030-39634-3_7.
[132] H. J. Pandit, P. Ryan, G. P. Krog, M. Crane, and R. Brennan, 'Towards a Semantic Specification for GDPR Data Breach Reporting', in *Legal Knowledge and Information Systems*, IOS Press, 2023, pp. 131–136. Accessed: Feb. 12, 2025. [Online]. Available: https://ebooks.iospress.nl/volumearticle/65578
[133] H. J. Pandit, 'A semantic specification for data protection impact assessments (dpia)', in *Towards a Knowledge-Aware AI*, IOS Press, 2022, pp. 36–50. Accessed: Feb. 12, 2025. [Online]. Available: https://ebooks.iospress.nl/volumearticle/60709
[134] 'ODRL Information Model 2.2'. Accessed: Feb. 12, 2025. [Online]. Available: https://www.w3.org/TR/odrl-model/
[135] E. Grunewald, P. Wille, F. Pallas, M. C. Borges, and M.-R. Ulbricht, 'TIRA: An OpenAPI Extension and Toolbox for GDPR Transparency in RESTful Architectures', presented at the International Workshop on Privacy Engineering (IWPE'21), 2021. doi: 10.1109/EuroSPW54576.2021.00039.
[136] E. Grünewald and F. Pallas, 'TILT: A GDPR-Aligned Transparency Information Language and Toolkit for Practical Privacy Engineering', in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event Canada: ACM, Mar. 2021, pp. 636–646. doi: 10.1145/3442188.3445925.
[137] R. Cole, S. Purao, M. Rossi, and M. Sein, 'Being Proactive: Where Action Research Meets Design Research'.
[138] L. A. Macaulay, *Requirements Engineering*. Springer Science & Business Media, 2012.
[139] M. Saunders, P. Lewis, and A. Thornhill, 'Research methods forbusiness students', *Essex: Prentice Hall: Financial Times*, 2003, Accessed: Jun. 28, 2025. [Online]. Available: https://www.researchgate.net/profile/Lysias-Charumbira/post/what_is_the_best_referenes_about_Research_Methodology_on_the_field_of_M

anagement/attachment/59d63d8a79197b807799a545/AS%3A420260026044416%401477209206490/download/Research_methods_for_business_students_f.pdf
[140] 'Data on the Web Best Practices'. Accessed: Aug. 23, 2025. [Online]. Available: https://www.w3.org/TR/dwbp/
[141] N. A. Krans, A. Ammar, P. Nymark, E. L. Willighagen, M. I. Bakker, and J. T. K. Quik, 'FAIR assessment tools: evaluating use and performance', *NanoImpact*, vol. 27, p. 100402, Jul. 2022, doi: 10.1016/j.impact.2022.100402.
[142] A. Cedergren and H. Hassel, 'Using Action Design Research for Developing and Implementing a Method for Risk Assessment and Continuity Management', *Safety Science*, vol. 151, p. 105727, Jul. 2022, doi: 10.1016/j.ssci.2022.105727.
[143] J. L. Barros-Justo, F. B. V. Benitti, and S. Tiwari, 'The impact of Use Cases in real-world software development projects: A systematic mapping study', Jun. 16, 2019, *arXiv*: arXiv:1906.06754. doi: 10.48550/arXiv.1906.06754.
[144] M. Sabou and M. Fernandez, 'Ontology (Network) Evaluation', in *Ontology Engineering in a Networked World*, M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, Eds., Berlin, Heidelberg: Springer, 2012, pp. 193–212. doi: 10.1007/978-3-642-24794-1_9.
[145] M. GRUNINGER, 'Methodology for the design and evaluation of ontologies', *Proc. IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995, Accessed: Jul. 26, 2025. [Online]. Available: https://cir.nii.ac.jp/crid/1571135650978573440
[146] M. Fernandez, A. Gomez-Pearez, and N. Juristo, 'Methontology: From Ontological Art Towards Ontological Engineering'.
[147] 'FOOPS!' Accessed: Jan. 21, 2025. [Online]. Available: https://catalogue.fair-impact.eu/resources/foops
[148] K. Supekar, 'A Peer-review Approach for Ontology Evaluation'.
[149] 'RDF 1.1 Concepts and Abstract Syntax'. Accessed: Jan. 12, 2025. [Online]. Available: https://www.w3.org/TR/rdf11-concepts/
[150] S. Cox, D. Browning, R. Albertoni, A. G. Beltran, P. Winstanley, and A. Perego, 'Data catalog vocabulary (DCAT) - version 2', W3C, W3C recommendation, Feb. 2020. [Online]. Available: https://www.w3.org/TR/vocab-dcat-2/
[151] 'What Is a Data Catalog? Importance, Benefits & Features | Alation'. Accessed: Jan. 12, 2025. [Online]. Available: https://www.alation.com/blog/what-is-a-data-catalog/
[152] P. Subramaniam, Y. Ma, C. Li, I. Mohanty, and R. C. Fernandez, 'Comprehensive and Comprehensible Data Catalogs: The What, Who, Where, When, Why, and How of Metadata Management', Feb. 01, 2023, *arXiv*: arXiv:2103.07532. doi: 10.48550/arXiv.2103.07532.
[153] 'EIF Perspective attributes of DCAT Application Profile for data portals in Europe | Interoperable Europe Portal'. Accessed: Jan. 12, 2025. [Online]. Available: https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/eif-perspective
[154] 'DCAT-AP 3.0'. Accessed: Jan. 12, 2025. [Online]. Available: https://semiceu.github.io/DCAT-AP/releases/3.0.0/#Dataset
[155] R. K. Yin, 'Case study research and applications'. Sage Thousand Oaks, CA, 2018. Accessed: Oct. 30, 2024. [Online]. Available: https://www.academia.edu/download/106905310/Artikel_Yustinus_Calvin_Gai_Mali.pdf
[156] 'TopQuadrant - TopBraid SPIN API'. Accessed: Jan. 17, 2025. [Online]. Available: https://www.topbraid.org/spin/api/
[157] 'OpenRefine'. Accessed: Jan. 15, 2025. [Online]. Available: https://openrefine.org/
[158] 'GraphDB Free Version 9.7.0-1 (Accessed: 2022-03-10)'. Accessed: May 08, 2022. [Online]. Available: https://graphdb.ontotext.com/
[159] D. F. Polit and C. T. Beck, 'Generalization in quantitative and qualitative research: Myths and strategies', *International Journal of Nursing Studies*, vol. 47, no. 11, pp. 1451–1458, Nov. 2010, doi: 10.1016/j.ijnurstu.2010.06.004.

[160] H. Kallio, A. Pietilä, M. Johnson, and M. Kangasniemi, 'Systematic methodological review: developing a framework for a qualitative semi-structured interview guide', *Journal of Advanced Nursing*, vol. 72, no. 12, pp. 2954–2965, Dec. 2016, doi: 10.1111/jan.13031.
[161] M. B. Miles, A. M. Huberman, and J. Saldaña, *Qualitative data analysis: a methods sourcebook*, Edition 3. Los Angeles London New Delhi Singapore Washington DC: Sage, 2014.
[162] 'Accountability Framework'. Accessed: Nov. 28, 2024. [Online]. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/accountability-framework/
[163] M. Bennett, 'The Zachman Framework Evolution by: John P. Zachman', Zachman International - FEAC Institute. Accessed: Aug. 23, 2025. [Online]. Available: https://zachman-feac.com/resources/ea-articles-reference/175-the-zachman-framework-evolution
[164] N. Moore, *How to Do Research: A Practical Guide to Designing and Managing Research Projects*. Facet Publishing, 2006.
[165] S. Liao *et al.*, 'Understanding GDPR Non-Compliance in Privacy Policies of Alexa Skills in European Marketplaces', in *Proceedings of the ACM Web Conference 2024*, Singapore Singapore: ACM, May 2024, pp. 1081–1091. doi: 10.1145/3589334.3645409.
[166] 'RDF Schema 1.1'. Accessed: Feb. 04, 2025. [Online]. Available: https://www.w3.org/TR/rdf-schema/
[167] *COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT REPORT Accompanying the document Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act)*. 2022. Accessed: Jan. 12, 2025. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022SC0721
[168] M. Wylot, M. Hauswirth, P. Cudré-Mauroux, and S. Sakr, 'RDF Data Storage and Query Processing Schemes: A Survey', *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, Jul. 2019, doi: 10.1145/3177850.
[169] A. Miles, B. Matthews, M. Wilson, and D. Brickley, 'SKOS Core: Simple knowledge organisation for the Web', in *Proceedings of the International Conference on Dublin Core and Metadata Applications*, Dublin Core Metadata Initiative, Sep. 2005. doi: 10.23106/dcmi.952107985.
[170] S. Auer, J. Lehmann, A.-C. N. Ngomo, and A. Zaveri, 'Introduction to Linked Data and Its Lifecycle on the Web', in *Reasoning Web*, vol. 8067, in Lecture Notes in Computer Science, vol. 8067. , Springer, 2013, pp. 1–90. doi: 10.1007/978-3-642-39784-4_1.
[171] 'SPARQL 1.1 Query Language'. Accessed: Jul. 29, 2025. [Online]. Available: https://www.w3.org/TR/sparql11-query/
[172] A. K. M. B. Haque *et al.*, 'Semantic Web in Healthcare: A Systematic Literature Review of Application, Research Gap, and Future Research Avenues', *Int J Clin Pract*, vol. 2022, p. 6807484, Oct. 2022, doi: 10.1155/2022/6807484.
[173] 'Collibra Data Catalog | Collibra'. Accessed: Jan. 12, 2025. [Online]. Available: http://localhost:4321/us/en/products/data-catalog.html
[174] 'DCAT-AP - data.gov.ie'. Accessed: Mar. 10, 2022. [Online]. Available: https://data.gov.ie/dataset/dcat-ap
[175] 'Software Requirements Specifications', IEEE Computer Society. Accessed: Feb. 02, 2025. [Online]. Available: https://www.computer.org/resources/software-requirements-specifications/
[176] 'Records of processing and lawful basis'. Accessed: Mar. 10, 2022. [Online]. Available: https://ico.org.uk/for-organisations/accountability-framework/records-of-processing-and-lawful-basis/
[177] 'Opinion 25/2022 regarding the European Privacy Seal (EuroPriSe) certification criteria for the certification of processing operations by processors | European Data Protection Board'. Accessed: Dec. 12, 2022. [Online]. Available: https://edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-252022-regarding-european-privacy-seal_en
[178] M. C. Suárez-Figueroa, A. Gómez-Pérez, and B. Villazón-Terrazas, 'How to Write and Use the Ontology Requirements Specification Document', in *On the Move to Meaningful Internet Systems:*

OTM 2009, vol. 5871, R. Meersman, T. Dillon, and P. Herrero, Eds., in Lecture Notes in Computer Science, vol. 5871. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 966–982. doi: 10.1007/978-3-642-05151-7_16.
[179] M. Horridge, 'A Practical Guide To Building OWL Ontologies Using Prot´eg´e 4 and CO-ODE Tools Edition 1.2', p. 109.
[180] 'How do we document our processing activities?' Accessed: Sep. 16, 2023. [Online]. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/documentation/how-do-we-document-our-processing-activities/
[181] 'content.pdf'. Accessed: Feb. 12, 2025. [Online]. Available: https://travesia.mcu.es/server/api/core/bitstreams/008888f7-05ff-4e12-bf18-7ba4f2e80182/content
[182] 'CKAN - The open source data management system (Accessed: 2022-03-10)', ckan.org. Accessed: May 08, 2022. [Online]. Available: http://ckan.org/
[183] 'FOAF Vocabulary Specification'. Accessed: Feb. 09, 2025. [Online]. Available: http://xmlns.com/foaf/spec/
[184] 'Apache Jena - Home'. Accessed: Jan. 17, 2025. [Online]. Available: https://jena.apache.org/
[185] 'Ontotext Refine', Ontotext. Accessed: Jan. 15, 2025. [Online]. Available: https://www.ontotext.com/products/ontotext-refine/
[186] P. Cupoli, S. Earley, and D. Henderson, 'DAMA-DMBOK2 Framework', 2012.
[187] G. Feretzakis, E. Vagena, K. Kalodanis, P. Peristera, D. Kalles, and A. Anastasiou, 'GDPR and Large Language Models: Technical and Legal Obstacles', *Future Internet*, vol. 17, no. 4, p. 151, Apr. 2025, doi: 10.3390/fi17040151.
[188] 'us-tackling-data-challenges-for-modernizing-legacy-technology-platforms.pdf'. Accessed: Feb. 12, 2025. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-tackling-data-challenges-for-modernizing-legacy-technology-platforms.pdf
[189] D. Golpayegani, H. J. Pandit, and D. Lewis, 'AICat: An AI Cataloguing Approach to Support the EU AI Act', Dec. 19, 2024, *arXiv*: arXiv:2501.04014. doi: 10.48550/arXiv.2501.04014.

Appendices

Appendix A Survey of Data Protection Professionals

Title: Evaluation of feature requirements for an Automated GDPR RoPA

Section 1 Overview

The GDPR requires organisations maintain a record of personal data processing activities (RoPA). Our research has shown that organisations are challenged with this activity. We have developed a prototype of an automated solution that enables any RoPA to be loaded, shared, exchanged, and examined between actors such as organisational units, processors, joint controllers, auditors, and regulators. This survey seeks to gather the opinions of data protection practitioners in the following areas:

- What are the biggest issues organisations are facing with RoPA?
- Who are the actors that would interoperate with RoPA?
- What are the highest priority areas that the automated RoPA could be used to address?
- What are the most important features that this automated RoPA system should contain to meet the needs of data protection practitioners?

Section 2 Plain Language Statement

Introduction to the Research Study: The GDPR requires organisations to maintain a record of personal data processing activities. Our research has shown that organisations are challenged by this activity. We have developed a prototype of an automatic RoPA, thus enabling it to be interoperable between actors. This survey seeks to establish what are the most important features that an automated RoPA system must contain. Participants are asked to complete this survey based on their practical data protection experience to date. All data collected will be anonymous and will be stored securely on DCU computer systems. There is no perceived risk to participants taking part in this research, but if any data subject has a concern, they can contact paul.ryan76@mail.dcu.ie. The research involves the School of Computing in Dublin City University. The principal investigators are Prof. Martin Crane, martin.crane@dcu.ie and Dr. Rob Brennan, rob.brennan@adaptcentre.ie. Some important notes that participants should be aware of:

- Information provided in this survey will be used for academic research and future academic publications. For updates on this research participants should contact paul.ryan76@mail.dcu.ie
- Participants are free not to be a part of this study and skip any question they wish.

- Participants have the right to withdraw from the study at any point. They only need to email using the contact details above to withdraw from the study.
- Participants do not have to supply any personal data to participate but if they do supply their name and/or email address, the information they share will be anonymised, which means that their name will not be used and what they say will not be directly attributable to them. There are no expected risks or benefits to them from participation.
- Information participants share will be anonymised and managed within the DCU data protection and freedom of information procedures. Anonymised information could be shared by the researcher in academic publications. It may be shared with relevant senior staff in DCU only if failure to do so was a breach of the law. Data will only be used for the purpose which is collected for and will be destroyed after the research purposes are fulfilled.
- The maximum data retention period for the anonymous data collected is 2024 upon which the data will be securely disposed of
- All data will be stored securely however participants need to be made aware that confidentiality of information provided cannot always be guaranteed by researchers and Confidentiality of information can only be protected within the limitations of the law - i.e., it is possible for data to be subject to subpoena, freedom of information claim or mandated reporting by some professions.
- Participants may withdraw from the Research Study at any point.
- You should explain to the participant that their participation in the project will end, at the point they withdraw, and refer to the data protection/privacy notice as to what will happen regarding their data. For example, withdrawing consent may mean that no future data collection will take place but previously collected data will still be processed etc.

If participants have concerns about this study and wish to contact an independent person, please contact: The Secretary, Dublin City University Research Ethics Committee, c/o Research and Innovation Support, Dublin City University, Dublin 9. Tel 01-7008000, e-mail rec@dcu.ie.

Section 3 Informed Participant Consent Form

In this section we need to gather your consent to ensure that you are fully informed about this research survey and ensure that you are happy to progress with completing the survey. The research title is 'Evaluation of feature requirements for a GDPR Automated Interoperable RoPA '. This survey seeks to establish what are the most important features that automated interoperable RoPA system must contain? The research involves the School of Computing in Dublin City University. The principal Investigators are Paul Ryan paul.ryan@mail.dcu.ie, Prof. Martin Crane martin.crane@dcu.ie and Dr. Rob Brennan, Rob.brennan@dcu.ie. All questions are optional, and the responses will be anonymous. To progress this research, we need your

consent to process your submission. Participants are required to confirm that they have read the Plain Language Statement and provide their informed consent to participate in the survey by ticking yes or no to the following questions.

- I have read the Plain Language Statement in the previous section.
- I understand the information provided in relation to data protection set out in the Plain Language Section
- I have had an opportunity to ask questions and discuss this study, should I wish to do so (contact paul.ryan@mail.dcu.ie)
- I have received satisfactory answers to all my questions.
- I confirm that I am aware that my involvement in the Research Study is voluntary, and I may withdraw from the Research Study at any point.
- I have read and understood the arrangements to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations.
- I have read and understood the arrangements for the retention and disposal of this research data.
- I have read and understood confirmations relating to any other relevant information as indicated in the Plain Language Section

- **Question 1:** I consent to participate in this study. **Response:** (Yes or No)

Section 4 Some general information about you?

Please note all responses are completely anonymous.

- **Question 2:** How many years have you worked in the Data Protection / Privacy domain? **Response:** multichoice answers (i) Less than one year (ii) 1 - 3 years (iii) Greater than three years (iv) Don't work in Data Protection / Privacy.
- **Question 3:** Do you hold or previously held the role of Data Protection Officer for an organisation? **Response:** multichoice answers (i) Yes (ii) No
- **Question 4:** List any of these disciplines where you hold a qualification? **Response:** Checkbox - Multiple answers allowed (i) Qualified Lawyer or Solicitor (ii) Information Technology / Computer Science Degree or greater (iii) Data protection certificate, diploma, degree or similar e.g. CIPP

Section 5 Your Organisation's RoPA

The RoPA is a critical GDPR compliance document. In this section, we try to evaluate the existing challenges that organisations are facing with the maintenance of RoPA. Please review the following statement and provide your opinion on their accuracy.

- **Question 5:** My organisation is challenged with keeping an accurate RoPA. **Response:** Likert scale – Strongly Disagree to Strongly Agree
- **Question 6:** Our RoPA is fully up to date. **Response:** Likert scale – Strongly Disagree to Strongly Agree
- **Question 7:** Our RoPA comprehensively describes our processing activities. **Response:** Likert scale – Strongly Disagree to Strongly Agree.
- **Question 8:** There is a lack of buy-in with RoPA within my organisation **Response:** Likert scale – Strongly Disagree to Strongly Agree

Section 6 Enablers of an Interoperable GDPR tools

Our Research has identified some key enablers that may need to be in place for automated interoperable GDPR tools to become established. By interoperable, we mean 'able to exchange and make use of information.' To what extent do you agree with these statements?

- **Question 9:** The adoption of developments in technologies would help with the automation of GDPR compliance. **Response:** Likert scale – Strongly Disagree to Strongly Agree
- **Question 10:** Automated interoperable GDPR tools will only happen if the Data Protection Supervisory Authorities drives adoption. **Response:** Likert scale – Strongly Disagree to Strongly Agree
- **Question 11:** For automated GDPR compliance tools to be developed, there will need to be agreement on common standards/ agreed semantics (definitions of terms) for personal data processing. **Response:** Likert scale – Strongly Disagree to Strongly Agree
- **Question 12:** A strong data governance platform within an organisation would enable the development of automated GDPR compliance tools. **Response:** Likert scale – Strongly Disagree to Strongly Agree

Section 7 An automated RoPA - how you use it and for what would you use it?

An automated RoPA system would provide the ability for Data Processing information to be automatically exchanged/shared and inspected between data processing actors. These actors could include internal organisational units/ departments, data processors, data controllers, auditors, regulators, or the data protection officer. These parties all have a requirement for access to, sharing or input to RoPA records.

- **Question 13:** If an automated RoPA was available, which actors' relationship would gain the most benefit from it, in your opinion? Please rank your top three in order. **Response:** (i) Intra organisation & DPO (ii) Processors & controller (iii) Joint controller & joint controller (iv) Controller & external DPO (v) Controller and regulator (vi) Controller & auditor

- **Question 14:** An automated RoPA could be used for many GDPR accountability tasks to assist the DPO. Can you select any areas where the automated RoPA might be used to address, in your opinion? **Response:** Checkbox (i) used to assist with supplier due diligence, (ii) checking of technical and organisational measures (iii) used for risk analysis of processes, (iv) to identify gaps, conflicts or non-compliances in RoPA, (v) to Identify where data transfers are occurring (vi) to ensure privacy notice is reflective, (vii) ability to verify data processing agreements are accurate, (viii) information to help with data subject rights requests, (ix) information to assist with data breaches, (x) information to help in the Data Protection Impact Assessment process.
- **Question 15:** For the automated RoPA, how important is it to that the solution has a search functionality to easily find information. Can you please rank on the scale below? **Response:** Likert Scale – Not Important to very important.
- **Question 16:** For the automated RoPA, how important is it that the solution is easy to use? Can you please rank on the scale below? **Response:** Likert Scale – Not Important to very important
- **Question 17:** For the automated RoPA, how important is it to that the solution is easy to implement. Can you please rank on the scale below? **Response:** Likert Scale – Not Important to very important
- **Question 18:** For the automated RoPA, how important is it to that the cost of investment is low. Can you please rank on the scale below? **Response:** Likert Scale – Not Important to very important
- **Question 19:** For the automated RoPA, how important is it to have a data quality check to ensure that the automated RoPA inputs is in place. How important is this feature, can you please rank on the scale below? **Response:** Likert Scale – Not Important to very important

Section 8 Thank you for your input to this research.

- **Question 20:** Do you have any other comments or observations you would like to add? **Response:** Long text answer

Appendix B Mapping of GDPR RoPA Concepts to DPV terms (2024)

The following table summarises the mapping between CSM-ROPA fields and DPV concepts. The column **'GDPR'** specifies relevant clauses 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 needed 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.

GDPR	Field	DPV	Map	DC	DP
5	Location of personal data	dpv:StorageLocation	Exact	R	R
5.1	Data Sources	dpv:DataSource	Exact	R	O
6.1	Legal Basis	dpv:LegalBasis	Exact	M	O
6.1	Link to record of consent	dpv:Consent	Exact	R	R
9.1	Special Personal Data Categories	dpv:SpecialCategoryPersonalData	Exact	R	O
9.1	Vulnerable Data Subjects	dpv:VulnerableDataSubject	Exact	R	O
22.1	Automated decision-making, or profiling	dpv:AutomatedDecisionMaking	Exact	R	R
26.1	Joint Controller Agreement	dpv:JointDataControllersAgreement	Exact	R	Not applicable
28	Data Processor	dpv:DataProcessor	Exact	R	M
28.3	Data Processing Contract	dpv:DataProcessingAgreement	Exact	R	R
28.3	Data Processor Contract	dpv:ControllerProcessorAgreement	Exact	R	R
30.1	Status of processing	dpv:Status	Exact	M	M
32	Technical/Organisational measures	dpv:Technology	Exact	R	R
32	Security Measures	dpv:TechnicalOrganisationalMeasure	Exact	R	R
32	Technologies used	dpv:Technology	Exact	R	R
33.5	Relevant Personal Data Breach	dpv:DataBreachRecord	Exact	R	R
35	Risk management	dpv:Risk, dpv:RiskMitigationMeasure	Exact	R	O
35	Relevant DPIA	dpv:DPIA	Exact	R	O
36.1	Impact Assessment	dpv:ImpactAssessment	Exact	R	R
35	DPIA Results	dpv:DPIA	Exact	R	R

GDPR	Field	DPV	Map	DC	DP
36.1	Prior Consultation	dpv:Consultation	Exact	R	R
37.6	External DPO organisation	dpv:DataProtectionOfficer	Exact	R	R
_	Business Process	dpv:Process	Exact	O	O
_	Process Owner	dct:contactPoint	Exact	M	M
_	Type of Processing	dpv:Processing	Exact	O	O
13/14/15	Data Subject Rights	dpv:DataSubjectRight	Exact	R	O
28/30.1(c)	Data Categories Transfer to third parties	dpv:Transfer, dpv:PersonalData	Exact	R	R
30(1)(a)	Data Controller Name	dpv:DataController	Exact	M	M
30(1)(a)	Data Controller Contact Details	dpv:hasName, dpv:hasContact	Exact	M	M
30(1)(a)	Data Protection Officer	dpv:DataProtectionOfficer	Exact	MC	MC
30.1(a)	Data Protection Officer Contact	dpv:hasName, dpv:hasContact	Exact	MC	MC
30.1(a)	Representative	dpv:Representative	Exact	MC	Not applicable
30.1(a)	Representative Contact	dpv:hasName, dpv:hasContact	Exact	MC	Not applicable
30.1(a)	Name of joint controller	dpv:JointDataController	Exact	MC	Not applicable
30.1(a)	Contact details of joint controller	dpv:hasName, dpv:hasContact	Exact	MC	Not applicable
30.1(b)	Purposes	dpv: Purpose	Exact	M	O
30.1(b)	Main/Auxiliary Processing	dpv:Importance	Exact	O	O
30.1(c)	Personal Data	dpv:PersonalDataCategory	Exact	M	M
30.1(c)	Data Subjects	dpv:DataSubject	Exact	M	M
30.1(d)	Recipients	dpv:Recipient	Exact	MC	MC
30.1(e)	Third countries for Transfers	dpv:ThirdCountry	Exact	MC	MC
30.1(e)	Safeguards	dpv:Safeguard	Exact	MC	MC
30.1(f)	Data Retention/Deletion Periods	dpv:StorageDuration,	Exact	M	O
30.1(g)	Technical/Organisational measures	dpv:TechnicalOrganisationalMeasure	Exact	M	M
44-47	Nature of Transfer	dpv:DataTransferLegalBasis	Exact	MC	MC
6.1(f)	Legitimate interests	dpv:LegitimateInterest	Exact	R	R
6.1(f)	Legitimate interests assessment	dpv:LegitimateInterestAssessment	Exact	R	R
6/14/30.1(b)	Data Combination	dpv:Combine	Exact	R	O

Appendix C CSM-RoPA 0.3 Classes and Relationships 2024

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
Personal Data Processing	dpv:PersonalDataHandling	Indicates association with Personal Data	dpv:hasPersonalDataHandling	dpv:ROPA	dpv:PersonalDataHandling	
Location of personal data	dpv:StorageLocation	Location or geospatial scope where the data is stored	dpv:hasStorage	dpv:PersonalDataHandling	dpv:StorageLocation	R
Data Sources	dpv:DataSource	The source or origin of the data	dpv:hasDataSource	dpv:PersonalDataHandling	dpv:DataSource	R
Legal Basis	dpv:LegalBasis	Legal basis used to justify processing of data or use of technology by a law	dpv:hasLegalBasis	dpv:PersonalDataHandling	dpv:LegalBasis	M
Record of consent	dpv:Consent	Consent of the Data Subject for specified process or activity	dpv:hasLegalBasis	dpv:PersonalDataHandling	dpv:Consent	R
Special Personal Data Categories	dpv:SpecialCategoryPersonalData	Sensitive Personal Data whose use requires specific additional legal permission or justification	rdfs:subClassOf	dpv:PersonalData	dpv:SpecialCategoryPersonaldata	R
Vulnerable Data Subjects	dpv:VulnerableDataSubject	Data Subjects which should be considered 'vulnerable' and, therefore, would require additional measures and safeguards	rdfs:subClassOf	dpv:DataSubject	dpv:VulnerableDataSubject	R
Automated decision-making or profiling	dpv:AutomatedDecisionMaking	Automated decision-making can be defined as ‘the ability to make decisions by technological means without human involvement.’ (‘Guidelines on Automated individual	dpv:hasContext	dpv:Processing	dpv:AutomatedDecision Making	R

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)', 2018, p. 8)				
Joint Controller Agreement	dpv:JointDataControllersAgreement	An agreement outlining conditions, criteria, obligations, responsibilities, and specifics for conducting the processing of data between Controllers within a Joint Controllers relationship	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:JointDataControllers Agreement	R
Data Processor	dpv:DataProcessor	A 'processor' means a natural or legal person, public authority, agency, or other body which processes data on behalf of the controller.	dpv:hasDataProcessor	dpv:PersonalDataHandling	dpv:LegalEntity	M
Data Processing Contract	dpv:DataProcessingAgreement	an agreement outlining conditions, criteria, obligations, responsibilities, and specifics for conducting the processing of data	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv: DataProcessingAgreement	R
Data Processor Contract	dpv:ControllerProcessorAgreement	Creation, completion, fulfilment, or performance of a contract, with the Data Controller and Data Processor as parties, and involving specified	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:ControllerProcessor Agreement	R

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		processing of data or technologies				
Status of processing	dpv:Status	The status or state of something	dpv:hasStatus	dpv:PersonalDataHandling	dpv:Status	R
Technical/Organisational measures	dpv:Technology	The technology, technological implementation, or any techniques, skills, methods, and processes used or applied	dpv:hasTechnicalOrganisational Measure	dpv:PersonalDataHandling	dpv:TechnicalOrganisational Measure	M
Security Measures	dpv:TechnicalOrganisational Measure	Technical and Organisational measures used to safeguard and ensure good practices in connection with data and technologies	dpv:hasTechnicalOrganisational Measure	dpv:PersonalDataHandling	dpv:TechnicalOrganisationa measure	R
Technologies used	dpv:Technology	The technology, technological implementation, or any techniques, skills, methods, and processes used or applied	dpv: isImpementedUsingTechnology	dpv: TechnicalOrganisational Measure	dpv:Technology	R
Relevant Personal Data Breach	dpv:DataBreachRecord	Record of a data breach incident	dpv:associatedWithDataBreach	dpv:PersonalDataHandling	dpv:DataBreachRecord	R
Risk management	dpv:Risk, dpv:RiskMitigationMeasure	Risks can be associated with one or more different concepts such as purpose, processing, personal data, technical or organisational measure	dpv:hasRisk, dpv:isMitigatedByMeasure	dpv:PersonalDataHandling	dpv:Risk, dpv:RiskMitigationMeasure	R

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
Relevant DPIA	dpv:DPIA	Impact assessment determines the potential and actual impact of processing activities on individuals or groups of individuals and considering the impacts of activities on their rights and freedoms	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:DPIA	R
Impact Assessment	dpv:ImpactAssessment	Calculating or determining the likelihood of impact of an existing or proposed process can involve risks or detriments.	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:ImpactAssessment	R
DPIA Results	dpv:DPIA	Impact assessment determines the potential and actual impact of processing activities on individuals or groups of individuals and considering the impacts of activities on their rights and freedoms	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:DPIA	R
Prior Consultation	dpv:Consultation	Consultation is a process of receiving feedback, advice, or opinion from an external agency	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:Consultation	R
External DPO organisation	dpv:DataProtectionOfficer	An entity within or authorised by an organisation to monitor internal compliance, inform, and advise on data protection	dpv:hasDataProtectionOfficer	dpv:PersonalDataHandling	dpv:DataProtectionOfficer	MC

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		obligations and function as a contact point for data subjects and the supervisory authority.				
Business Process	dpv:Process	An action, activity, or method	dpv:hasPersonalDataHandling	dpv:PersonalDataHandling	dpv:Process	R
Process Owner	dct:contactPoint	Contact details of the entity	dcat:contactPoint	dpv:PersonalDataHandling	vcard:Kind	R
Type of Processing	dpv:Processing	Operations or 'processing' performed on data	dpv:hasProcessing	dpv:PersonalDataHandling	dpv:Processing	R
Data Subject Rights	dpv:DataSubjectRight	The rights applicable or provided to a Data Subject	dpv:hasRight	dpv:Data Subject	dpv:DataSubjectRight	R
Data Categories Transfer to third parties	dpv:Transfer, dpv:PersonalData	The personal data categories that are transferred to third-party recipients	dpv:hasRecipient	dpv:PersonalDataHandling	dpv:Transfer	R
Data Controller Name	dpv:DataController	The individual or organisation that decides (or controls) the purpose(s) of processing personal data.	dpv:hasDataController	dpv:PersonalDataHandling	dpv:DataController	MC
Data Controller Contact Details	dpv:hasName, dpv:hasContact	Contact details of the entity	dpv:hasName; dpv:hasContact	dpv:DataController	rdfs:Literal	M
Data Protection Officer	dpv:DataProtectionOfficer	An entity within or authorised by an organisation to monitor internal compliance, inform, and advise on data protection obligations and function as a contact point for data	dpv:hasDataProtectionOfficer	dpv:DataController	dpv:DataProtectionOfficer	MC

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		subjects and the supervisory authority.				
Data Protection Officer Contact	dpv:hasName, dpv:hasContact	Contact details of the entity	dpv:hasName; dpv:hasContact	dpv:DataProtectionOfficer	rdfs:Literal	MC
Representative	dpv:Representative	A representative of a legal entity	dpv:hasRepresentative	dpv:DataController	dpv:Representative	MC
Representative Contact	dpv:hasName, dpv:hasContact	Contact details of the entity	dpv:hasName; dpv:hasContact	dpv:Representative	rdfs:Literal	MC
Name of joint controller	dpv:JointDataController	A group of Data Controllers that jointly determine the purposes and means of processing	dpv:hasJointDataController	dpv:DataController	dpv:JointDataController	MC
Contact details of joint controller	dpv:hasName, dpv:hasContact	Contact details of the entity	dpv:hasName, dpv:hasContact	dpv:JointDataController	rdfs:Literal	MC
Purposes	dpv: Purpose	The purpose or goal here is intended to sufficiently describe the intention or objective of why the data or technology is being used and should be broader than mere technical descriptions of achieving a capability.	dpv:hasPurpose	dpv:PersonalDataHandling	dpv:Purpose	M
Main/Auxiliary Processing	dpv:Importance	Importance can be used to express importance, desirability, relevance, or significance as a context	dpv:hasContext	dpv:PersonalDataHandling	dpv:Importance	R
Personal Data	dpv:PersonalDataCategory	Types of information relating to an identified or identifiable natural	dpv:hasPersonalData	dpv:PersonalDataHandling	dpv:PersonalData	M

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		person ('data subject'); an identifiable natural person is one who can be identified				
Data Subjects	dpv:DataSubject	The individual (or category of individuals) whose personal data is being processed	dpv:hasDataSubject	dpv:PersonalDataHandling	dpv:DataSubject	M
Recipients	dpv:Recipient	Recipients that receive personal data can be a Third Party, Data Controller, or Data Processor.	dpv:hasRecipient	dpv:PersonalDataHandling	dpv:LegalEntity	MC
Third countries for Transfers	dpv:ThirdCountry	Indicates applicability or relevance of a 'third country'	dpv:hasThirdCountry	dpv:Transfer	dpv:ThirdCountry	MC
Safeguards	dpv:Safeguard	A safeguard is a precautionary measure for the protection against or mitigation of negative effects	dpv:hasOrganisationalMeasure	dpv:PersonalDataHandling	dpv:Safeguard	MC
Data Retention/Deletion Periods	dpv:StorageDuration,	Duration or temporal limitation on storage of data	dpv:hasStorage	dpv:PersonalDataHandling	dpv:StorageDuration	M
Technical/Organisational measures	dpv:TechnicalOrganisational Measure	Technical and Organisational measures used to safeguard and ensure good practices in connection with data and technologies	dpv:hasTechnicalOrganisational Measure	dpv:PersonalDataHandling	dpv:TechnicalOrganisational Measure	M
Nature of Transfer	dpv:DataTransferLegalBasis	Specific or special categories and instances	dpv:hasLegalBasis	dpv:Transfer	dpv:LegalBasis	MC

GDPR Concept	Ontology Term	Definition	Relation	Domain	Range	Nec.
		of legal basis intended for justifying data transfers				
Legitimate interests	dpv:LegitimateInterest	Legitimate Interests of the Data Subject in conducting specified activities	dpv:hasLegalBasis	dpv:PersonalDataHandling	dpv:LegitimateInterest	R
Legitimate interests assessment	dpv:LegitimateInterestAssessment	Indicates an assessment regarding the use of legitimate interest as a lawful basis by the data controller	dpv:hasOrganisationalMeasure	dpv: LegitimateInterest	dpv:LegitimateInterest Assessment	R
Data Combination	dpv:Combine	To join or merge data	rdfs:subClassOf	dpv:Combine	dpv:Processing	R

Appendix D Serialised RDF representation of Data Protection Officer Class.

```
dpv:DataProtectionOfficer a rdfs:Class,
    skos:Concept ;
  dct:contributor "Georg P. Krog, Paul Ryan" ;
  dct:created "2020-11-04"^^xsd:date ;
  dct:modified "2021-12-08"^^xsd:date ;
  dct:source [ a schema:WebPage ;
      schema:name "GDPR Art.37" ;
      schema:url "https://eur-lex.europa.eu/eli/reg/2016/679/art_37/oj" ] ;
  rdfs:isDefinedBy dpv: ;
  rdfs:subClassOf dpv:Representative ;
  sw:term_status "accepted"@en ;
  skos:broader dpv:Representative ;
  skos:definition "An entity within or authorised by an organisation to monitor internal compliance, inform and advise on data protection obligations and act as a contact point for data subjects and the supervisory authority."@en ;
  skos:inScheme dpv:entities-legalrole-classes ;
  skos:prefLabel "Data Protection Officer"@en.
```

Appendix E Serialised RDF representation of Object Property.

```
dpv:hasDataProtectionOfficer a rdf:Property,
    skos:Concept ;
  dcam:rangeIncludes dpv:DataProtectionOfficer ;
  dct:contributor "Paul Ryan, Rob Brennan" ;
  dct:created "2022-03-02"^^xsd:date ;
  rdfs:isDefinedBy dpv: ;
  rdfs:subPropertyOf dpv:hasRepresentative ;
  sw:term_status "accepted"@en ;   skos:broader dpv:hasRepresentative ;
  skos:definition "Specifies an associated data protection officer"@en ;
```

Appendix F How Competence Questions are met using terms from CSM-RoPA.

Question No.	Competency Questions	CSM-RoPA Class	CSM-RoPA Property
CQ01	What is the purpose of the data processing activity?	dpv:Purpose	dpv:hasPurpose
CQ02	What categories of personal data are being processed?	dpv:PersonalDataCategory	dpv:hasPersonalData
CQ03	Who is the data controller responsible for the processing activity?	dpv:DataController	dpv:hasDataController
CQ04	Who are the data processors involved in the processing activity?	dpv:DataProcessor	dpv:hasDataProcessor
CQ05	What is the legal basis for the processing (e.g., consent, contract, legal obligation)?	dpv:LegalBasis	dpv:hasLegalBasis
CQ06	What is the duration or retention period for the data being processed?	dpv:StorageDuration,	dpv:hasStorage
CQ07	Who are the data subjects whose personal data is being processed?	dpv:DataSubject	dpv:hasDataSubject
CQ08	What are the data subjects' rights in this processing activity?	dpv:DataSubjectRight	dpv:hasRight
CQ09	Are there any third parties or recipients with whom the data is shared?	dpv:Transfer, dpv:PersonalData, dpv:Recipient	dpv:hasRecipient
CQ10	Is the personal data transferred to countries outside the European Economic Area (EEA)?	dpv:Transfer, dpv:ThirdCountry	dpv:hasRecipient
CQ11	What safeguards are in place for data transfers to third countries?	dpv:Safeguard	dpv:hasOrganisationalMeasure
CQ12	What security measures are applied to protect personal data during processing?	dpv:TechnicalOrganisationalMeasure	dpv:hasTechnicalOrganisationalMeasure
CQ13	What are the risks associated with the data processing activities?	dpv:Risk	dpv:hasRisk, dpv:isMitigatedByMeasure
CQ14	Has a Data Protection Impact Assessment (DPIA) been conducted for this processing activity?	dpv:DPIA	dpv:hasOrganisationalMeasure
CQ15	What is the source of the data being processed?	dpv:DataSource	dpv:hasDataSource
CQ16	Are any automated decision-making processes involved in this data-processing activity?	dpv:AutomatedDecisionMaking	dpv:hasContext
CQ17	What procedures are in place for data deletion once the retention period expires?	dpv:StorageDuration, dpv:TechnicalOrganisationalMeasure	dpv:hasStorage
CQ18	Who is the Data Protection Officer (DPO) overseeing this processing activity?	dpv:DataProtectionOfficer	dpv:hasDataProtectionOfficer
CQ19	What technical and organisational measures of security are in place?	dpv:TechnicalOrganisationalMeasure	dpv:hasTechnicalOrganisationalMeasure
CQ20	What are the contact details of the data controller's representative?	_	dpv:hasName, dpv:hasContact
CQ21	Who is the representative of the data controller?	dpv:Representative	dpv:hasRepresentative
CQ22	What is the nature of the personal data transfer?	dpv:DataTransferLegalBasis	dpv:hasLegalBasis

Question No.	Competency Questions	CSM-RoPA Class	CSM-RoPA Property
CQ23	Who are the Joint Controllers involved in the processing activity?	dpv:JointDataController	dpv:hasJointDataController
CQ24	What are the contact details of the joint controller?	_	dpv:hasName, dpv:hasContact
CQ25	What are the contact details of the data protection officer?	_	dpv:hasName, dpv:hasContact
CQ26	What are the contact details of the data controller?	_	dpv:hasName, dpv:hasContact
CQ27	Where is the data processing contract located?	dpv:DataProcessingAgreement	dpv:hasOrganisationalMeasure
CQ28	What is the name or identifier of the business process?	dpv:Process	dpv:hasPersonalDataHandling
CQ29	Who is the owner of the processing activity?	dct:contactPoint	dcat:contactPoint
CQ30	What categories of special personal data are processed in the processing activity?	dpv:SpecialCategoryPersonalData	_
CQ31	What is the status of the processing activity?	dpv:Status	dpv:hasStatus
CQ32	What categories of vulnerable data subjects are processed in the processing activity?	dpv:VulnerableDataSubject	_
CQ33	What is the type of processing?	dpv:Processing	dpv:hasProcessing
CQ34	Where is the record of consent located?	dpv:Consent	dpv:hasLegalBasis
CQ35	Where is the joint controller agreement located?	dpv:JointDataControllersAgreement	dpv:hasOrganisationalMeasure
CQ36	Has there been an impact assessment/ prior consultation for this processing activity?	dpv:ImpactAssessment	dpv:hasOrganisationalMeasure
CQ37	Has a personal data breach occurred related to this processing activity?	dpv:DataBreachRecord	dpv:associatedWithDataBreach
CQ38	What System or software is used (technologies used)?	dpv:Technology	dpv: isImpementedUsingTechnology
CQ39	What technical and organisational measures of security are in place?	dpv:TechnicalOrganisationalMeasure	dpv:hasTechnicalOrganisationalMeasure
CQ40	Where is the personal data located (to support data subject rights requests)?	dpv:StorageLocation	dpv:hasStorage
CQ41	Is this processing activity the organisation's main or auxiliary processing activity (distinguishing between main/core and auxiliary/secondary operations)?	dpv:Importance	dpv:hasContext
CQ42	Where is the Legitimate interest assessment located?	dpv:LegitimateInterest	dpv:hasOrganisational Measure
CQ43	Who is the external entity acting as the organisation's DPO?	dpv:DataProtectionOfficer	dpv:hasDataProtectionOfficer
CQ44	Have multiple data sets been combined for the processing activity?	dpv:Combine	_

Appendix G SPARQL Query to Represent RoPA with CSM-RoPA

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dpv: <http://www.w3.org/ns/dpv#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
select DISTINCT ?department ?title ?purpose ?purpose_category ?datasubject ?personaldata ?recipient
?recipient_category ?recipient_location ?storageyears ?measures ?period_start ?period_end ?department
?contactname ?contactemail where {

#   dataset metadata
   ?dataset a dcat:Dataset .
   ?dataset dct:title ?title .

#   processing period of operation
   ?dataset dct:temporal ?period .
   ?period dct:startDate ?period_start .
   ?period dct:endDate ?period_end .

#   processing point of contact
   ?dataset dct:creator/foaf:name ?department .
   ?dataset dcat:contactPoint/foaf:name ?contactname .
   ?dataset dcat:contactPoint/dpv:hasContact ?contactemail .

##   processing metadata
#   purpose
   ?dataset dpv:hasPurpose/dct:title ?purpose .
   ?dataset dpv:hasPurpose/rdf:type ?purpose_category_i .
   ?purpose_category_i rdfs:subClassOf* dpv:Purpose .
   ?purpose_category_i rdfs:label ?purpose_category .
#   data subject
   ?dataset dpv:hasDataSubject/dct:title ?datasubject .
#   personal data
   ?dataset dpv:hasPersonalDataCategory ?personaldata_i .
   ?personaldata_i rdfs:subClassOf* dpv:PersonalDataCategory .
   ?personaldata_i rdfs:label|dct:title ?personaldata .
#   duration
   ?dataset dpv:hasDuration ?duration .
   ?duration a dpv:StorageDuration .
   { ?duration time:years ?storageyears . } UNION { FILTER NOT EXISTS { ?duration time:years ?storageyears
} . ?duration dct:title ?storageyears }
   ?dataset dpv:hasTechnicalOrganisationalMeasures/dct:title ?measures .
#   optional recipients
   OPTIONAL {
      ?dataset dpv:hasRecipient ?recipiententity .
      ?recipiententity dct:title ?recipient .
      ?recipiententity a/dct:title|a/rdfs:label ?recipient_category .
      OPTIONAL { ?recipiententity dpv:hasLocation ?recipient_location }
   }

} ORDER BY ?department ?title
```

Appendix H Sample Extracts taken from EDPS RoPA

A sample of EDPS RoPA Record 01a is provided below, refer to online resource[62] for all records utilised in case study.

EDPS record of processing activity

Record of EDPS activities processing personal data, based on Article 31 of Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC ('the Regulation').

Nr.	Item	Description
	Selection of staff for the EDPS Secretariat and the EDPB Secretariat	
1.	Last update of this record	**03/11/2021**
2.	Reference number	**01a**
	Part 1 - Article 31 Record (specific legal obligation to publish – see Article 31(5)) < row to be deleted when filled in>	
3.	Name and contact details of controller	European Data Protection Supervisor (EDPS) **Postal address**: Rue Wiertz 60, B-1047 Brussels **Office address**: Rue Montoyer 30, B-1000 Brussels **Telephone:** +32 2 283 19 00 **Email:** edps@edps.europa.eu **Website:** https://edps.europa.eu. Responsible service/unit: Human Resources, Budget, Administration (HRBA) Unit, Contacts: edps-selections@edps.europa.eu Contact form for enquiries on processing of personal data to be preferably used: https://edps.europa.eu/form/edpsweb-contact-form_en
4.	Name and contact details of DPO	edps-dpo@edps.europa.eu
5.	Name and contact details of joint controller (where applicable)	n/a
6.	Name and contact details of processor (where applicable)	n/a
7.	Very short description and purpose of the processing	Select staff for the EDPS Secretariat and the EDPB Secretariat. For general info on selection and recruitment at the EDPS: https://edps.europa.eu/careers_en The legal basis of the procedure is: -the Staff Regulations (and particularly Art. 27-34) and the Conditions of Employment of Other Servants of the EU -the Decision of the EDPS of 4 November 2020 adopting general implementing provisions relating to the engagement and the use of contract staff.
8.	Description of categories of persons whose data the EDPS processes and list of data categories	We process the following categories of personal data contained in the application of every person - candidate who sent his/her application to the functional mailbox (edps-selections@edps.europa.eu) (staff selection): - Data identifying the applicant and contact details (name, first name, gender, nationality, date and place of birth, postal and e-mail address, telephone number, mobile telephone number, fax number). - Data derived from the candidate's application and CV, his/her motivation letter and other supporting documents submitted, namely current entity of assignment or current employer EUI in case of inter-institutional and for external applicants: institution/company and department, country of residence, as well as function group, grade, step, seniority in the current job, type of post of the person, type of post of the current job, in case of AST officials applying for AD vacancies: information related to certification procedure - Documents requested in the vacancy notice to verify whether the application is admissible or not: a curriculum vitae, a covering letter or motivation letters and other supporting documents submitted by the applicants including information on education, competencies and language skills, diplomas and certificates,

[62] https://doi.org/10.5281/zenodo.14914848

Appendix I Mapping of ICO concepts to DPV terms

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
1	Organisation	DataController	Exact Match
4	Personal data	Categories of personal data/ categories of data subjects	Exact Match
6	Data map (including data flow and data audit)	Nil - ADD DATA MAP to CSM ROPA	DPV Under consideration
7	Up to date	Other Vocabulary	Other vocabulary
9	staff	Categories of data subjects	Exact Match
12	Processing activities	Type of Processing	Exact Match
13	Questionnaire / staff surveys	Other Vocabulary	Other vocabulary
15	Record of Processing Activities	Record of Processing Activities	Exact Match
16	Electronic form	technology	partial match
17	Information (regarding processing)	Record of Processing Activities	partial match
20	Policies and procedures	Technical and organisational measures of security	partial match
21	Accurate (of ROPA)	Record of Processing Activities	partial match
25	Types of data	Categories of personal data/ categories of data subjects	Exact Match
26	Data minimisation purposes	Technical and organisational measures of security	partial match
27	ROPA	Record of Processing Activities	Exact Match
28	Organisation's name	Organisation	Exact Match
29	Contact details	Other Vocabulary	Other Vocabulary
31	A Processor	Processor	Exact Match
32	The joint controller	Joint Controller	Exact Match
33	Their representative	Representative	Exact Match
34	The DPO	Data Protection Officer	Exact Match
35	Purposes of processing	Purposes of processing	Exact Match
36	A description of the categories of individuals and personal data;	Categories of personal data/ categories of data subjects	Exact Match
37	The categories of recipients of personal data	Categories of recipients of transfer data	Exact Match
38	Transfers to third countries	Third countries that personal data are transferred to	Exact Match

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
39	A record	Record of Processing Activities	Exact Match
40	The transfer mechanism safeguards	Safeguard for data Transfer	Exact Match
41	Retention schedules	Retention/Deletion Periods	Exact Match
42	Technical and organisational security measures	Technical and organisational measures of security	Exact Match
43	An internal record of all processing activities	Record of Processing Activities	Exact Match
44	Any processors	Processor	Exact Match
47	Links to documentation	Location of personal data	Exact Match
48	Information required for privacy notices	Privacy Notice	Exact Match
49	The lawful basis for the processing	Legal Basis for Processing	Exact Match
50	The source of the personal data	The original source of data	Exact Match
51	Records of consent	Link to record of consent	Exact Match
52	Controller-processor contracts	Legal Agreement	Exact Match
53	The location of personal data	Location of personal data	Exact Match
54	DPIA reports .	Data Protection Impact Assessment	Exact Match
55	Records of personal data breaches	Personal Data Breach	Exact Match
56	Information required for processing special category data or criminal conviction and offence data	Special Category Personal Data	Exact Match
57	Data Protection Act 2018 (DPA 2018)	Nil - add Data Protection Regulation to CSM ROPA	DPV Under consideration
58	Retention and erasure policy documents.	Retention and erasure policy.	Exact Match
61	Each activity	Type of Processing	Exact Match
62	A review DATE	Other Vocabulary	Other vocabulary
65	Reasons FOR LGAL BASES CHOSEN	Legal Basis for Processing	partial match
67	Special category data	Special Category Personal Data	Exact Match
68	Criminal offence data	Criminal	Exact Match
69	A lawful basis	Legal Basis for Processing	Exact Match
70	General processing	Type of Processing	partial match
71	An additional condition for processing this type of data	Type of Processing	partial match
73	The official authority	Nil - ADD Data Protection Regulator to CSM ROPA	Exact Match

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
75	Criminal offence data	Criminal	Exact Match
76	Consideration of the requirements of Article 9 or 10 of the GDPR	Special Category Personal Data	Exact Match
77	Schedule 1 of the DPA 2018 Schedule 1	Special Category Personal Data	partial match
78	An appropriate policy document	Technical and organisational measures of security	partial match
79	Schedule 1 conditions	duplicate	partial match
80	Compliance with the data protection principle relates to policies and procedures	technical and organisational measures	partial match
81	Special category or criminal offence data	Special Category Personal Data	Exact Match
82	Retention and erasure purposes	Retention and erasure policy.	Exact Match
85	The processing.	Type of Processing	Exact Match
86	Lawful basis	Legal Basis for Processing	Exact Match
87	New processing	Lawful basis; Data Protection Impact Assessment	complex match
89	The purposes of the processing	Purposes of processing	Exact Match
91	Relevant conditions for processing	Purposes of processing	partial match
92	Any special category data	Special Category Personal Data	Exact Match
94	Organisation's privacy notice(s).	Privacy Notice	Exact Match
96	Understandable format. (privacy notice)	Privacy notice	Exact Match
97	Change in circumstances	Risk - Information about the risk	complex match
99	New and unanticipated purpose	Purposes of processing	Exact Match
101	Timely manner	Other Vocabulary	Other vocabulary
102	the changes	Risk - Information about the risk / ROPA / DPIA	complex match
103	Consent requests	Consent record	Exact Match
104	Terms and conditions (Consent separate from other terms and conditions)	Legal Basis for Processing; consent; consent notice	partial match
105	A positive opt-in (description of consent process)	consent notice; has provision method	complex match
106	Use pre-ticked boxes (consent)	has consent notice; has provision method;	partial match
107	A pre-condition (description of consent process)	properties of consent process; is explicit; has consent notice	complex match
108	A service	Type of Processing	Exact Match
110	Consent	Legal Basis for Processing	Exact Match

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
111	Easy way;	properties of consent process (withdrawl of consent)	Exact Match
113	Names of any third parties (relying on consent)	name of controller ; legal basis ; consent	complex match
114	Consent	Legal Basis for Processing	Exact Match
115	Records	Link to record of consent	Exact Match
116	An individual	Data Subject	Exact Match
117	Consented to	Legal Basis for Processing	Exact Match
118	Consented.	Legal Basis for Processing	Exact Match
120	Relevant staff (ease of staff to access consent records)	controller; consent; is explicit	partial match
121	Evidence of consent	Consent	Exact Match
125	Example of consent	Consent	Exact Match
126	Online forms	technology	partial match
127	Notices	Privacy Notice / Location of data	Exact Match
128	Opt in-tick boxes (evidence of consent)	Privacy Notice	partial match
129	Paper-based forms.	technology	partial match
130	Procedure (check consents)	Technical and organisational measures of security; Consent; has expiry	complex match
131	Consents	Legal Basis for Processing	Exact Match
132	Relationship (check purposes of data use not changes)	Risk - Information about the risk ; Data Protection impact assessment ; technical and Organisational measures	partial match
133	Processing	Type of Processing	Exact Match
137	Consent	Legal Basis for Processing	Exact Match
138	Appropriate intervals.	Other Vocabulary	Other vocabulary
140	Privacy dashboards	technology	partial match
141	Preference-management tools	technology ; consent ; has withdrawl method	complex match
142	People	Categories of data subjects ; data subject	Exact Match
145	Reasonable efforts	property of policy	partial match
146	age	age verification	Exact Match
149	Child	Child	Exact Match
150	Reasonable and effective procedure	Technical and organisational measures of security	partial match
153	Record parental	location of personal data; legal basis ; consent	complex match

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
154	guardian consent.	Legal Basis for Processing has provision by justification	Exact Match
155	Online services	technology	partial match
158	Risk-based age-checking systems	Risk - Information about the risk / legal basis	complex match
160	Level of certainty	Risk - Information about the risk	complex match
161	risks	Risk - Information about the risk	Exact Match
162	children's rights and freedoms.	Data Subject Rights	Exact Match
163	online services	technology	partial match
165	child	Child	Exact Match
166	under thirteen	Other Vocabulary	Other vocabulary
167	records of parental or guardian consent	Link to record of consent	Exact Match
168	Reasonable efforts	property of verification process	complex match
170	parental or guardian responsibility.	has Provision by justification	Exact Match
171	particular consideration	Consent/ technology used / Vulnerable Data Subject Category	partial match
172	a child	Child	Exact Match
175	Legitimate Interests assessment	Legitimate Interest Assessment	Exact Match
176	Legitimate Interest	Legitimate interests for the processing	Exact Match
177	benefits of the processing	Legitimate interests for the processing ; impact assessment	partial match
178	A 'balancing test'	Legitimate interests for the processing ; impact assessment	partial match
181	people's data	Categories of personal data/ categories of data subjects	complex match
182	Intrusive ways	Risk - Information about the risk; purpose; consent ; Data Protection Impact Assessment	complex match
183	harm	Risk	Exact Match
184	a very good reason;	legitimate interests assessment	partial match
185	Interests of vulnerable groups	Vulnerable Data Subject Category	partial match
186	people with learning disabilities or children;	Vulnerable Data Subject Category	partial match
187	Safeguards	Appropriate Safeguards for Third Country Transfers,	Exact Match
188	negative impact	Risk - Information about the risk ; Data Protection Impact Assessment	Exact Match
189	An opt-out (part of LIA)	Opting out of process	Exact Match

Mapping internal number	Unique term from accountability framework to map to CSM ROPA	Matched Concept from CSM ROPA	Status of match
190	A DPIA	Data Protection Impact Assessment	Exact Match
191	Decision	Decision making	Exact Match
192	Assessment	Assessment	Exact Match
193	Start of processing	Status of processing	Other vocabulary
194	Outcome of LIA	Legitimate Interest Assessment	Exact Match

Appendix J Sample RDF Listing Generated from OntoRefine

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dpv: <https://w3id.org/dpv/dpv-skos#> .
@prefix dpv-gdpr: <https://w3id.org/dpv/dpv-skos/dpv-gdpr#> .
@prefix dpv-pd: <https://w3id.org/dpv/dpv-skos/dpv-pd#> .
@prefix dpcat: <https://w3id.org/dpcat#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix time: <https://www.w3.org/TR/owl-time/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dpv-legal: <https://www.w3id.org/dpv/dpv-skos/dpv-legal#> .
@prefix Upsilon: <https://w3id.org/dpcat/examples/Upsilon/vocab#> .
@prefix : <https://w3id.org/dpcat/examples/Upsilon/14#> .

: a dpcat:ROPA ;
   dct:title ‘Customer Service Procesing Activities ‘@en ;
   dct:description ‘Customer Service Processing Activities’@en ;
   skos:editorialNote ‘Upsilon have a customer service to deal with customer related service issues.’@en ;
   dct:created ‘2023-05-15’^^xsd:date ;
   dct:identifier ‘14’^^xsd:string ;
   dct:publisher Upsilon:UpsilonPLC ;
   dcat:contactPoint Upsilon:UpsilonPLC ;
            dpv:DataProtectionOfficer ‘Hilary Smith;’
            dcat:dataset :15-1.

:15-1 a dpcat:ROPARecord ;
   dct:title ‘ Financial Record Keeping ‘@en ;
   dct:description ‘ The organisations gathers transaction details of customer invoices. The data is stored on
the SAP ERP system  ‘@en ;
 skos:editorialNote ‘The data is retained to meet legal oblilgations of record keeping.  ‘@en ;
   dct:created ‘2023-05-17’^^xsd:date ;
   dct:publisher Upsilon:UpsilonPLC ;
   dcat:contactPoint Upsilon:Finance ;
   dpv:hasDataController Upsilon:UpsilonPLC ;
   dpv:hasResponsibleEntity Upsilon:Finance  ;
   dpv:hasPurpose dpv:RecordManagement, dpv:FulfilmentOfObligation, dpv:PaymentManagement ;
   dpv:hasDataSubject Upsilon:Customer ;
   dpv:hasDataSource dpv:DataSubject ;
   dpv:hasLegalBasis dpv:LegalObligation ;
            dpv:hasAutomatedDecisionMaking ‘No’;
   dpv:hasPersonalData dpv-pd:Name, dpv-pd:Purchase ,  dpv-pd:AccountIdentifier ;
   dpv:hasStorage [
      a dpv:StorageCondition ;
      dpv:hasDuration [
         a dpv:StorageDuration, time:Duration ;
         dct:description ‘The data collected is kept as a rule for a maximum of seven years. Once the legal
deadline has expired, the data is deleted.’@en ;
         time:numericDuration ‘7’^^xsd:decimal ;
         time:unitType :unitYear ;
         ] ] ;
   dpv:hasRecipient Upsilon:SalesForce  ;
   dpv:hasTechnicalOrganisationalMeasure Upsilon:StandardMeasures ;
   dpv:hasRight Upsilon:StandardRights  .
```

Appendix K DPCat RDF file prepared by Third Party Processor

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dpv: <https://w3id.org/dpv/dpv-skos#> .
@prefix dpv-gdpr: <https://w3id.org/dpv/dpv-skos/dpv-gdpr#> .
@prefix dpv-pd: <https://w3id.org/dpv/dpv-skos/dpv-pd#> .
@prefix dpcat: <https://w3id.org/dpcat#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix time: <https://www.w3.org/TR/owl-time/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dpv-legal: <https://www.w3id.org/dpv/dpv-skos/dpv-legal#> .
@prefix Upsilon: <https://w3id.org/dpcat/examples/Upsilon/vocab#> .
@prefix : <https://w3id.org/dpcat/examples/Upsilon/26#> .

: a dpcat:ROPA ;
   dct:title ‘Cloud Software Service for Recruitment  ‘@en ;
   dct:description ‘Cloud Software Service for Recruitment ‘@en ;
   skos:editorialNote ‘‘‘Cloud Software Service for Recruitment.’@en ;
   dct:created ‘2023-11-07’^^xsd:date ;
   dct:identifier ‘26’^^xsd:string ;
   dct:publisher Upsilon:SAP ;
   dcat:contactPoint Upsilon:SAP ;
            dpv:DataProtectionOfficer ‘Jane Doe ‘;
            dcat:dataset :26-1.

:26-1 a dpcat:ROPARecord ;
   dct:title ‘ Cloud Software Service for Recruitment ‘@en ;
   dct:description ‘ Cloud Software Service for Recruitment  ‘@en ;
 skos:editorialNote ‘‘‘The data is retained to meet legal oblilgations of record keeping.  ‘‘‘@en ;
   dct:created ‘2023-11-07’^^xsd:date ;
   dct:publisher Upsilon:SAP ;
   dcat:contactPoint Upsilon:SAP ;
   dpv:hasDataController Upsilon:SAP;
   dpv:hasResponsibleEntity Upsilon:SAP ;
   dpv:hasPurpose dpv:RecordManagement ;
   dpv:hasDataSubject Upsilon:Customer ;
   dpv:hasDataSource dpv:DataSubject ;
            dpv:hasAutomatedDecisionMaking ‘YES’;
   dpv:hasPersonalData dpv-pd:Name ;
   dpv:hasStorage [
      a dpv:StorageCondition ;
      dpv:hasDuration [
         a dpv:StorageDuration, time:Duration ;
         dct:description ‘The data collected is kept as a rule for a maximum of seven years. Once the legal
deadline has expired, the data is deleted.’@en ;
         time:numericDuration ‘7’^^xsd:decimal ;
         time:unitType :unitYear ;
         ] ] ;
   dpv:hasRecipient Upsilon:SAP  ;
   dpv:hasTechnicalOrganisationalMeasure Upsilon:StandardMeasures ;
   dpv:hasRight Upsilon:StandardRights  .
```

Appendix L Semi-Structured Interviews

Five data protection experts were interviewed between November and December 2024. The following are summarised notes, critical statements, and comments from the interviews.

Expert 1: The ERoPA Approach eliminates a lot of manual work, which makes it very useful. Using ERoPA for validation versus other documents is good, particularly if you can triangulate.
The advantage of the ERoPA Approach is that it will lead to a more granular ROPA.
How does the ERoPA Approach work from a blank canvas, i.e. no RoPA exist?
The ERoPA Approach will save work. The idea of validating RoPA data against other data sources is a good idea. Organisations are already thinking (this way) towards automating compliance, e.g., the generation of a privacy notice generated from ROPA.

Expert 2: I agree that DPOs are greatly challenged. We must bear in that ropa is a 'high-level crib sheet '. It does not have all the details of all GDPR compliance data.
Exchanging RoPA information with processors and org units is a good idea. This should lead to better accuracy and more up-to-date ROPA.
The idea of scheduling ROPA reviews automatically is good.
When comparing your ROPA with a data processing agreement or a privacy notice, you must be careful, as the language is different (depending on the audience). You may not be comparing like with like. Similarly, the auto-generation of privacy notice from RoPA may lack subtlety.
Technical organisational measures (GDPR Article 32) on RoPA can be quite high-level—the ROPA is really the centre of accountability and is quite often only a pointer to the relevant documents.
If you are to machine-read documents, you need to be rock solid on the 'meaning' of each term.
There may be a use case in vendor assessment where questions are asked, and yes/ no answers are required. We need to make sure the system is fully validated and tested . Regulator support would be beneficial.

Expert 3: Everyone expresses challenges when dealing with an Excel-based ROPA.
The ERoPA is only as good as the last update of the spreadsheet. So many contact points require a contribution to ROPA. DPOs will like this approach to RoPA.
Using Excel for RoPA is complex and excruciating. Reviewing RoPA line by line is hard. However, we agree that compliance gaps can be identified using EROPA.
In traditional approaches, documents are live and then they are out of date the day after. ERoPA may help to keep documents up to date.
The challenge in the DPO world is that compliance documents are not tracked, linked, and connected. The use of common semantics here is important to help compliance but we need to be careful of the semantics of the document's meaning.
ERoPA could be used to help manage risk. There could be a use case for Supplier Due diligence.

Expert 4: The interoperability of ROPA is a great idea, but it needs agreement on terms.
The support of the regulator in getting acceptance of ERoPA would be good.
Knowledge graphs in ROPA are very good, as the RoPA references many other documents.
The interoperability of information for RoPA is important.
Gathering RoPA information from the relevant personnel in a form suitable for RoPA can be challenging. A comprehensive ROPA is essential. Aligning privacy notices, data processing agreements, and ROPA is a long way off within organisations. The use of a knowledge graph approach to RoPA facilitates the interconnected nature of processing (e.g. IT systems/recipients/data employed) . The deep dive of RoPAs by regulators has confirmed that ROPAs are too shallow and need more detail.

Expert 5: A lot of RoPAs are Excel-based – any automation would help.

It can be difficult to check different documents for GDPR compliance, and there may be a lack of consistency across them.
The interoperability of ROPA between stakeholders is a great idea.
Common semantics are essential for interoperability.
The exchange of RoPA information between stakeholders is good for compliance verification.
Exchanging RoPA may not overcome the issue of lack of buy-in among stakeholders; however, it may work well with processors as buy-in can be forced on vendors/processors.
The ERoPA interoperability offers advantages for processors as they can align compliance with the controller and reduce expensive audits and manual data exchange and review.
The key thing with intra-organisation compliance verification is to make it easy for the stakeholders.
Using other documents to verify ROPA would be excellent and would work well.
the Interconnected nature of RoPA.