Engaging Content
Engaging People

Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV

Paul Ryan[123] , Harshvardhan J.Pandit[14] , Rob Brennan[12]

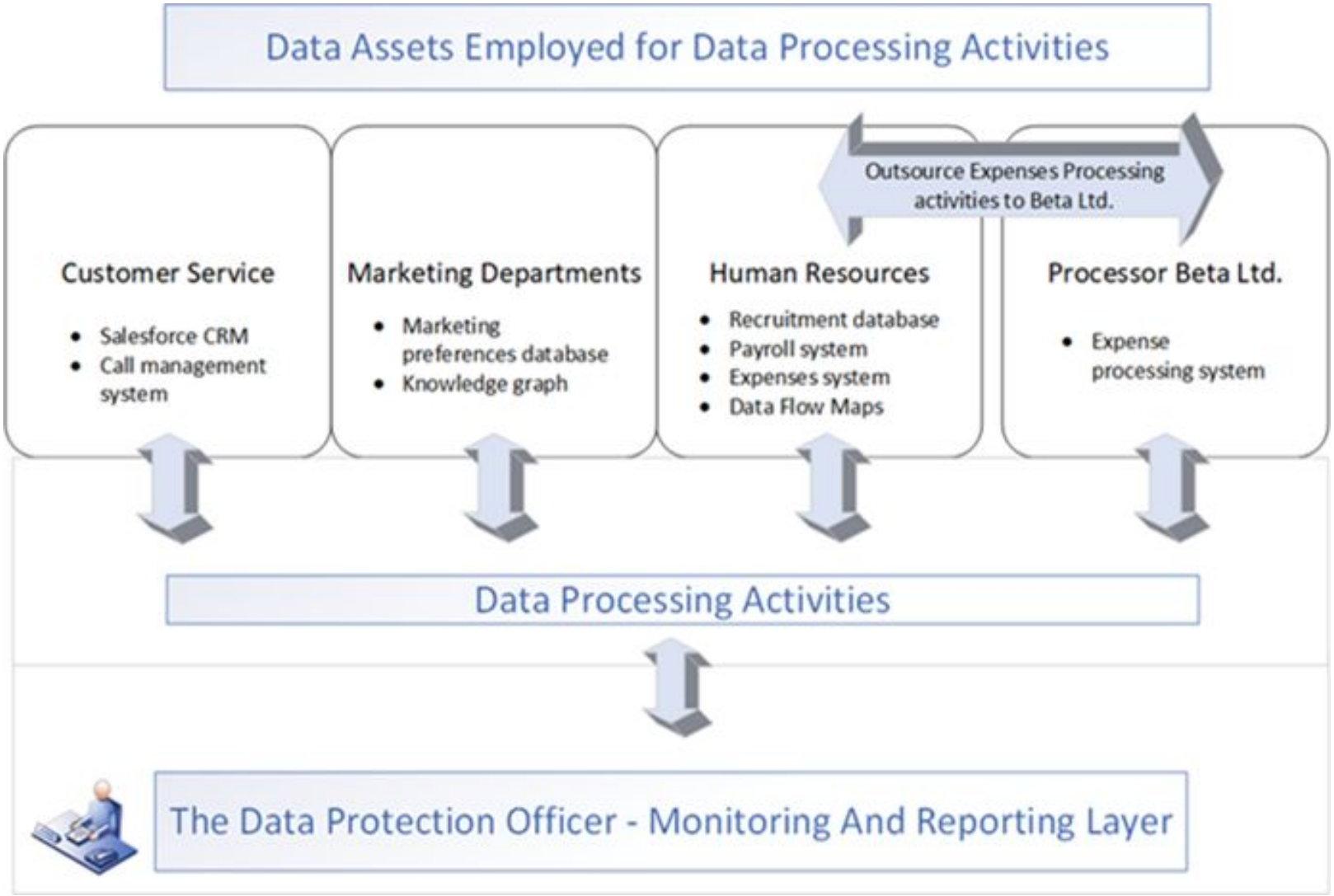
1. ADAPT Centre,
2. School of Computing, Dublin City University
3. Uniphar Plc, Ireland
4. Trinity College Dublin

Contact : Paul.Ryan76@mail.dcu.ie

Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union
European Union
European Regional Development Fund
SFI

The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Use case – Alpha Ltd.



"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Please Feel Sorry for the DPO

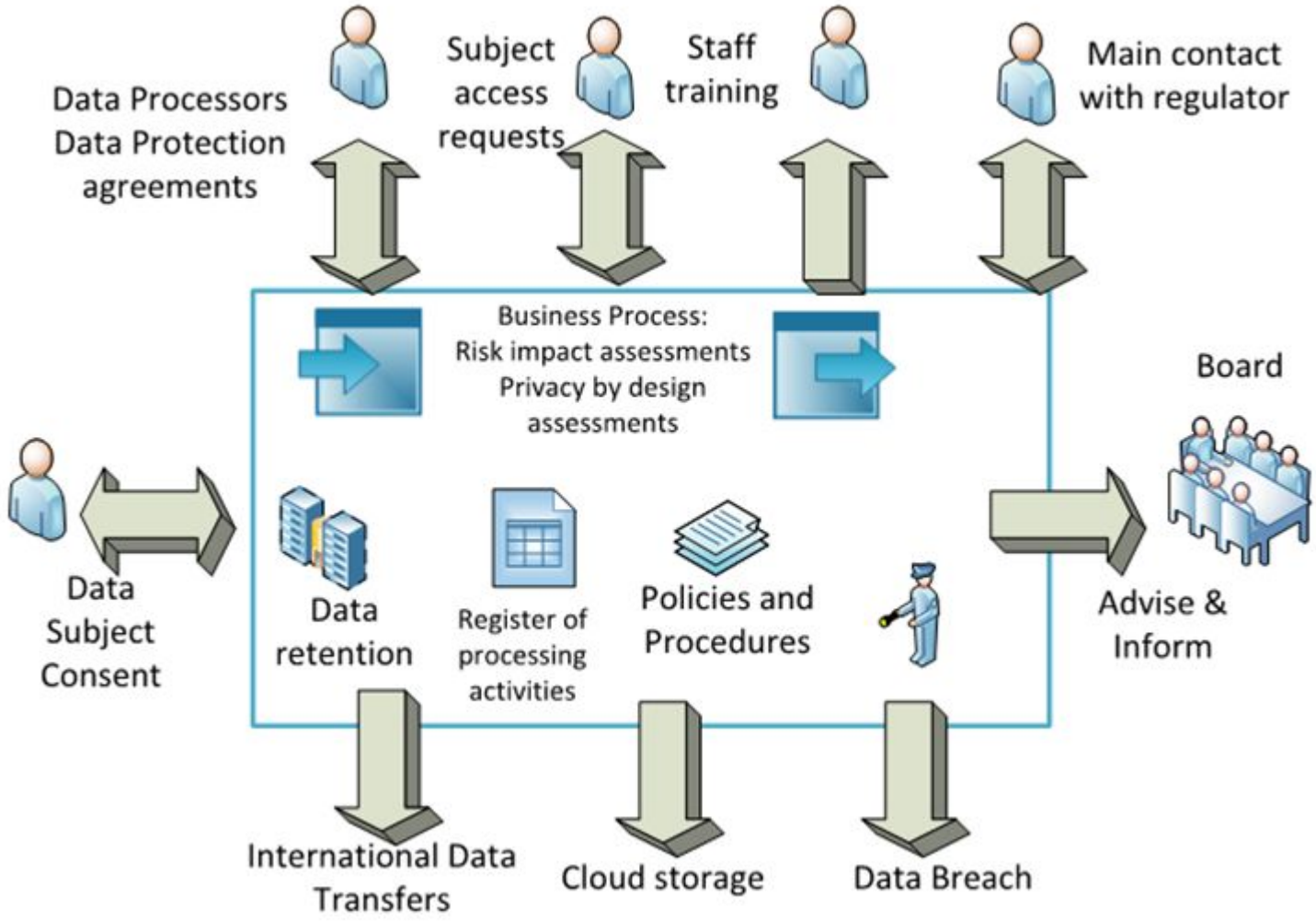


Image: Paul Ryan, Martin Crane, Rob Brennan, Design Challenges for GDPR Regtech, International Conference on Enterprise Information Systems, ICEIS 2020, http://doras.dcu.ie/24547/

Unfortunately that is a lot of Excel files…

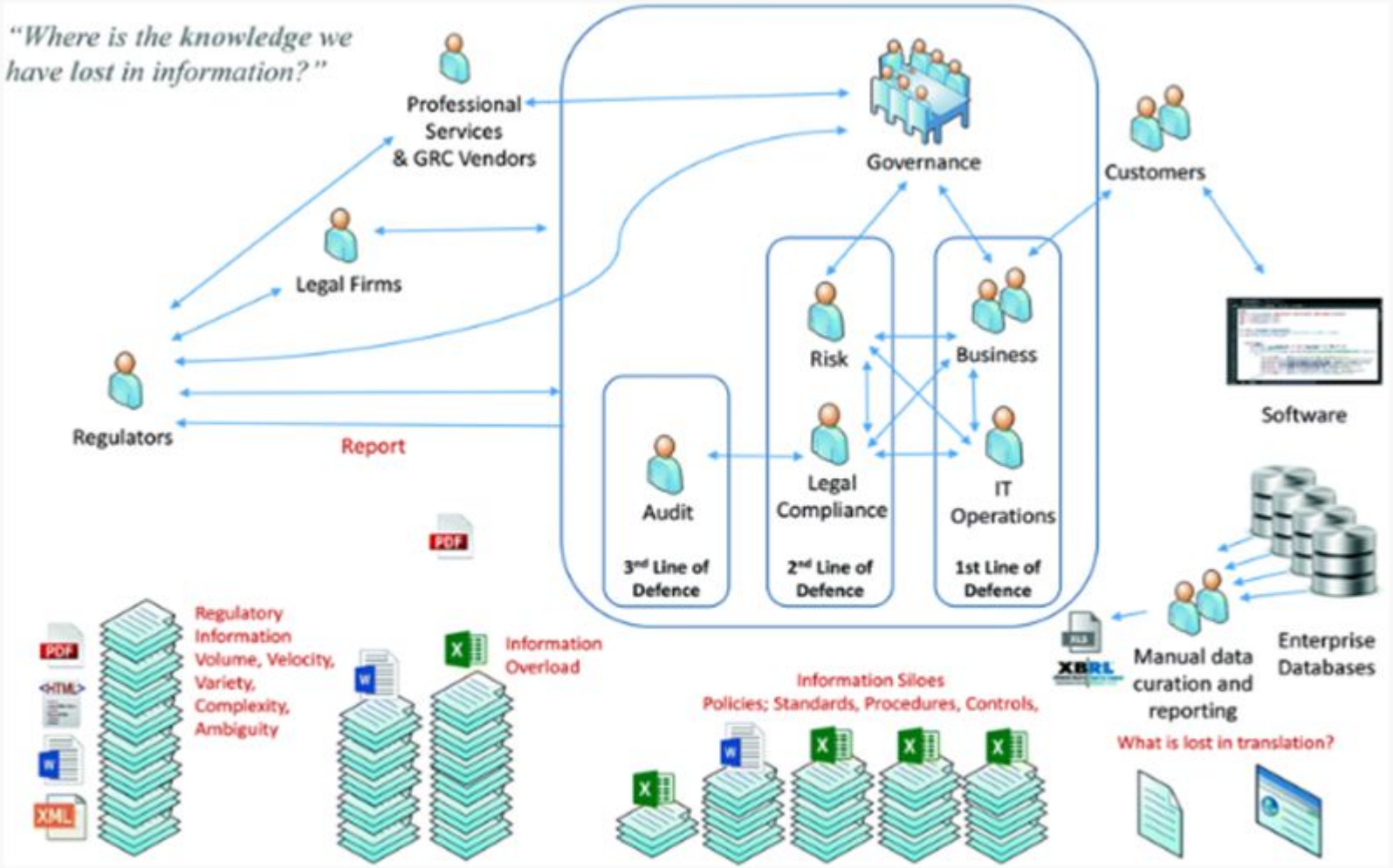


Fig. 6.1
Information overload, complexity, silos, and loss

Image: Butler T., O'Brien L. (2019) Understanding RegTech for Digital Regulatory Compliance. In: Lynn T., Mooney J., Rosati P., Cummins M. (eds) Disrupting Finance. Palgrave Studies in Digital Business & Enabling Technologies. Palgrave Pivot, Cham. https://doi.org/10.1007/978-3-030-02330-0_6

Motivation

- Organisations must maintain a register of processing activities (ROPA) to meet the accountability principle of the GDPR
- These data processing activity descriptions must be gathered from heterogeneous organisational sources such as departments, divisions, and external processors
- Many organisations already have diverse data collection tools for documenting data processing activities, and this heterogeneity is likely to grow in the future
- Most GDPR knowledge graph research to date has focused on Knowledge Graph representation and inference issues rather than integration and usability /deployment

"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

State of the Art

How do organisations capture and express data processing activities

- Commercial solutions - through informal tools, such as visual data flow mapping
- Customised in house software, and spreadsheets - stand-alone and lack interoperability , not sufficiently detailed, not kept up to date
- Enterprise Architecture - may not extend to entire organisation & Specialist tools required
- Many semantic–based projects provide vocabularies, ontologies, and policy languages to reference GDPR . They focus on legal compliance but don't consider how data is maintained or generated within/by organisations and the entities involved in this process
- Several semantic vocabularies exist but none have modelled a ROPA

"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Privacy Aware Data Governance

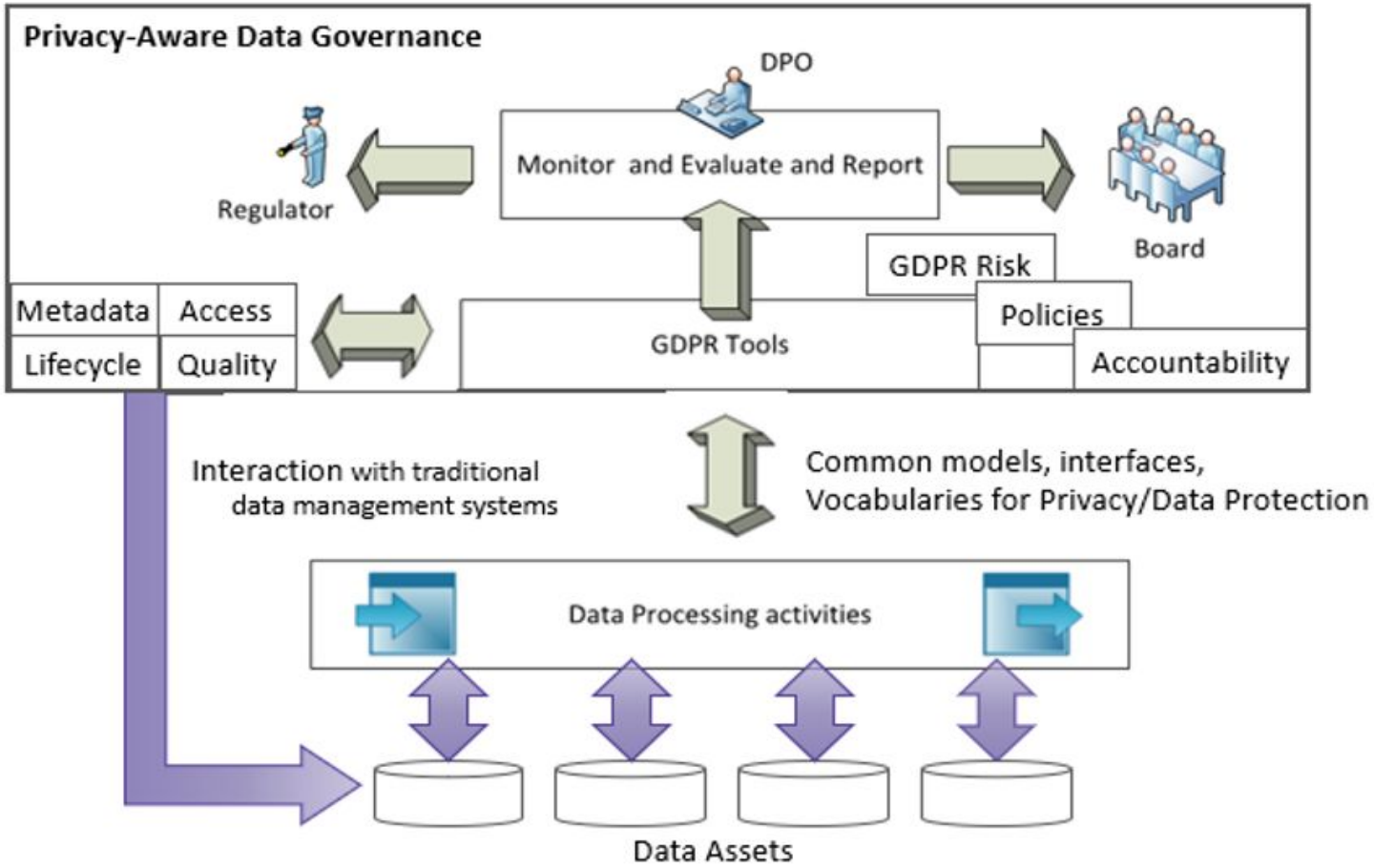


Image: Ryan, Paul and Brennan, Rob (2021) Demonstrating GDPR accountability with CSM-ROPA: extensions to the data privacy vocabulary. In: 24th International Conference Enterprise Information Systems (ICEIS '21), 26-28 Apr 2021 http://doras.dcu.ie/25797/

"Building a
Paul Ryan, H

A New Approach

- A new approach extending the well-known DCAT-AP[1] standard and utilising Data Privacy Vocabulary (DPV[2]) to express concepts necessary to complete a ROPA.

- This approach enables data catalog implementations to merge and federate the ROPA metadata for compliance related activities without requiring full alignment or merging of all the underlying data sources describing data processing activities.

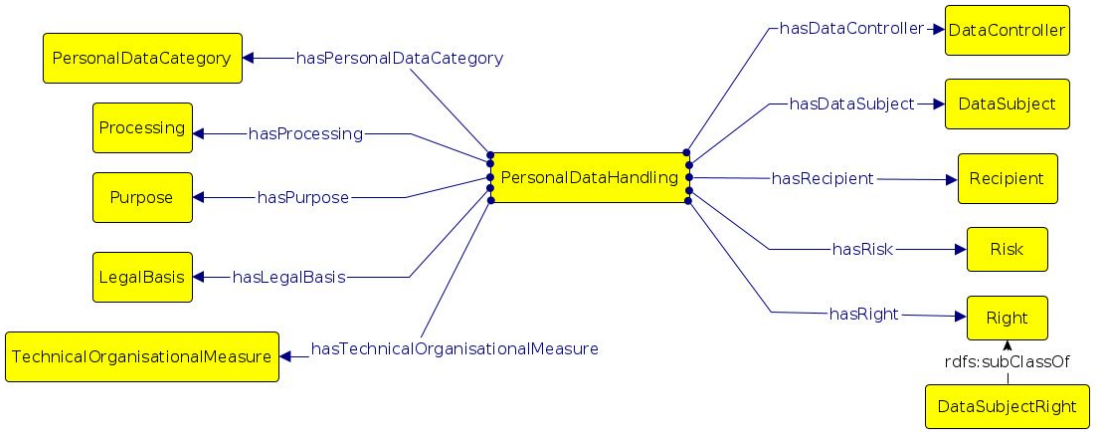
1 https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en
2 https://w3.org/ns/dpv

Our Research Goal

To establish the extent that a Data Processing Activities Catalog based on DCAT-AP and the Data Privacy Vocabulary (DPV) can overcome the heterogeneity of sources to generate and maintain a ROPA

What is the Data Privacy Vocabulary ?

- Semantic-Web vocabulary (terms) and ontology (relationships) of concepts associated with privacy and data protection, primarily derived from GDPR
- Enables automation of tools such as generating policies, reasoning, linking documentation, compliance assessments and evaluations
- A community specification by W3C Data Privacy Vocabulary and Controls Community Group (DPVCG).



"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Requirements for a tool to overcome the heterogeneity of sources to generate and maintain a ROPA

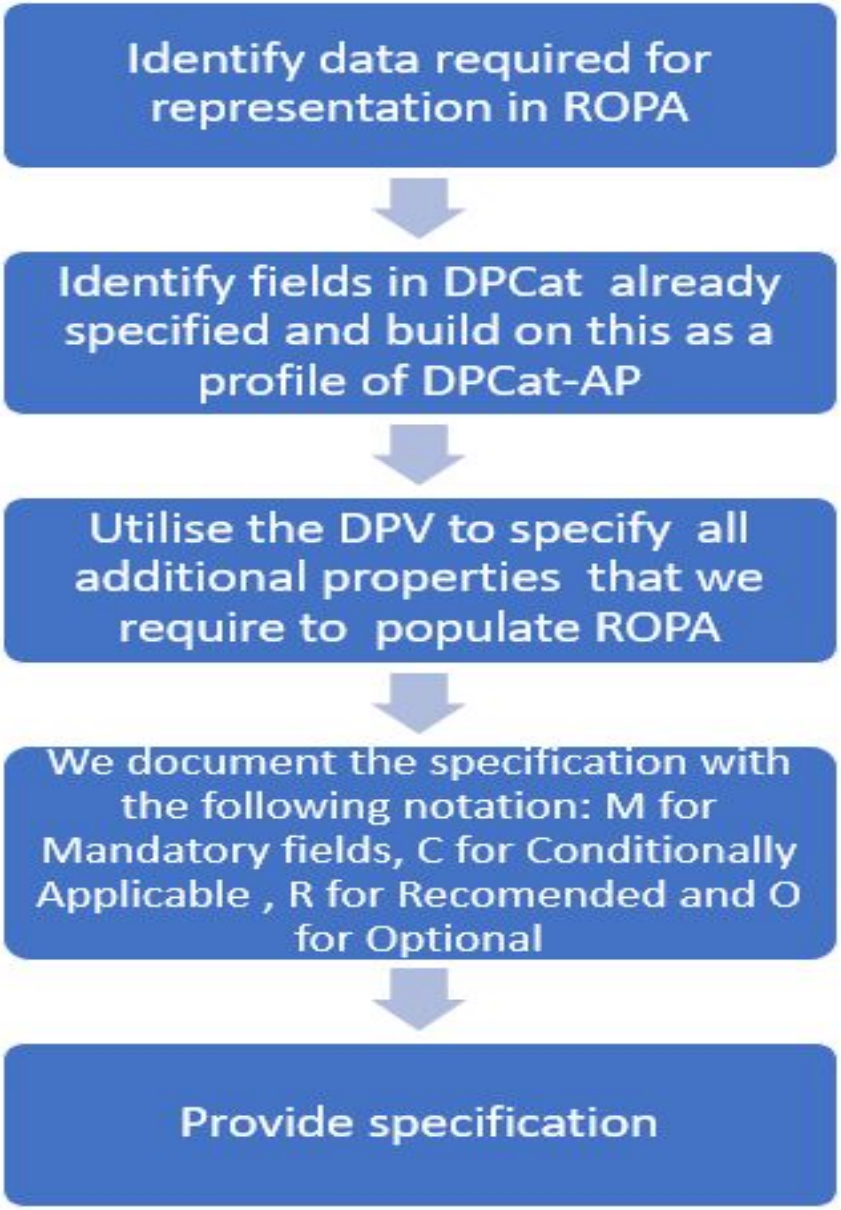
System Requirement	Basis for Requirement
Supports the heterogeneity of data sources	ICO accountability Tracker[1] Section 6.3.2 (all sources)
Enable standards-based collation of the data required for completion of a ROPA	ICO accountability Tracker[1] Section 6.3
Record temporal validity of processing activities	ROPA template[2] (Belgium)
Support periodic or continuous changes to data processing activity	ICO accountability Tracker[1] Section 6.1.2
Record identity of activity host and organisational unit and relevant contact	GDPR Art. 30.1 (a) Controller contact data
Facilitate searching records, e.g. identify activities active on a specific date	GDPR Art 37 DPO - monitor, advise & Inform
Enable the creation of ROPA and other compliance-related documentation using information collected in the records	GDPR Art. 24 Obligations of Controller, Art. 30 Register of processing activities)
Minimise the data to be collected and integrated	GDPR Compliance Tools – Best Practice from RegTech[3]
Easy to deploy, e.g. based on established or commonly used software platforms	GDPR Compliance Tools – Best Practice from RegTech[3]

1 https://ico.org.uk/media/for-organisations/documents/2618229/accountability-tracker.xlsx

2 https://www.gegevensbeschermingsautoriteit.be/professioneel/eerstehulp-avg/toolbox

3 Ryan P., Crane M., Brennan R. (2021) GDPR Compliance Tools: Best Practice from RegTech. In: Filipe J., Śmiałek M., Brodsky A., Hammoudi S. (eds) 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

DPCat Research Process



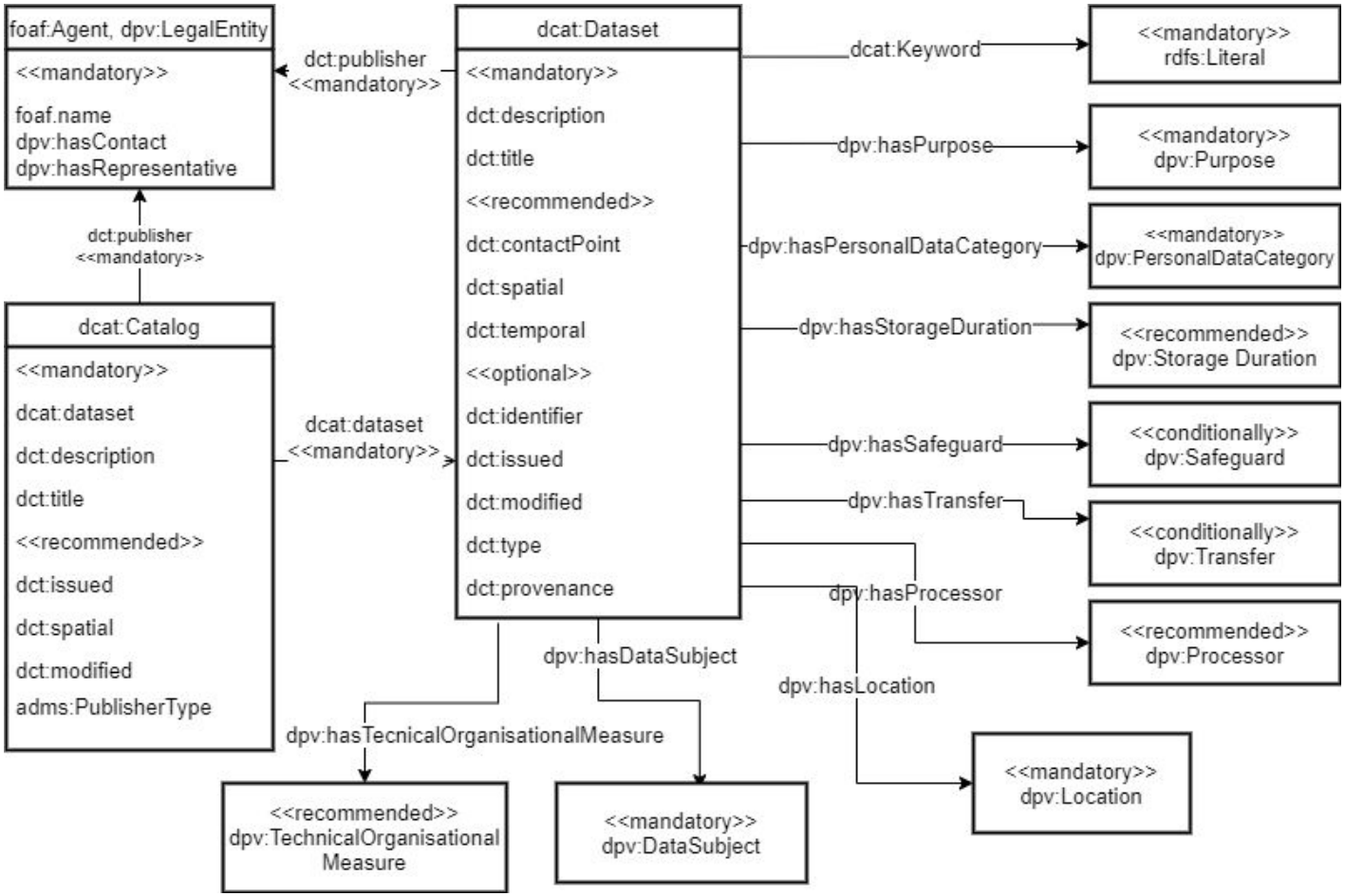
DPCat Specification

Specification for Representing the Data Processing Activities in DPCat

ROPA Requirement	GDPR Article 30 Obligation	DPCat Property	DPCat Property Range
Controller	Mandatory	dct:publisher	foaf:Agent, dpv:Controller, adms:PublisherType
Purpose	Mandatory	dpv:hasPurpose	dpv:Purpose
Categories of Data Subjects	Mandatory	dpv:hasDataSubject	subclass of dpv:DataSubject
Categories of Personal Data	Mandatory	dpv:hasPersonalDataCategory	subclass of dpv:PersonalDataCategory
Categories of Recipients	Conditionally Applicable	dpv:hasRecipient	subclass of foaf:Agent, adms:PublisherType, dpv:LegalEntity
Data Transfer	Conditionally Applicable	dpv:hasProcessing	dpv:Transfer
Data Transfer Location	Mandatory	dpv:hasLocation	dpv:Location
Data Transfer Recipient	Mandatory	dpv:hasRecipient	foaf:Agent, adms:PublisherType, dpv:LegalEntity
Data Transfer Safeguards (see note)	Conditionally Applicable	dpv:hasSafeguard	dpv:Safeguard
Time limits for erasure of different categories of data	Recommended	dpv:hasDuration	dpv:StorageDuration
Technical and Organisational Measures	Recommended	dpv:hasTechnicalOrganisationalMeasure	dpv:TechnicalOrganisationalMeasure
Processors responsible for processing	Recommended	dpv:hasRecipient	dpv:Processor

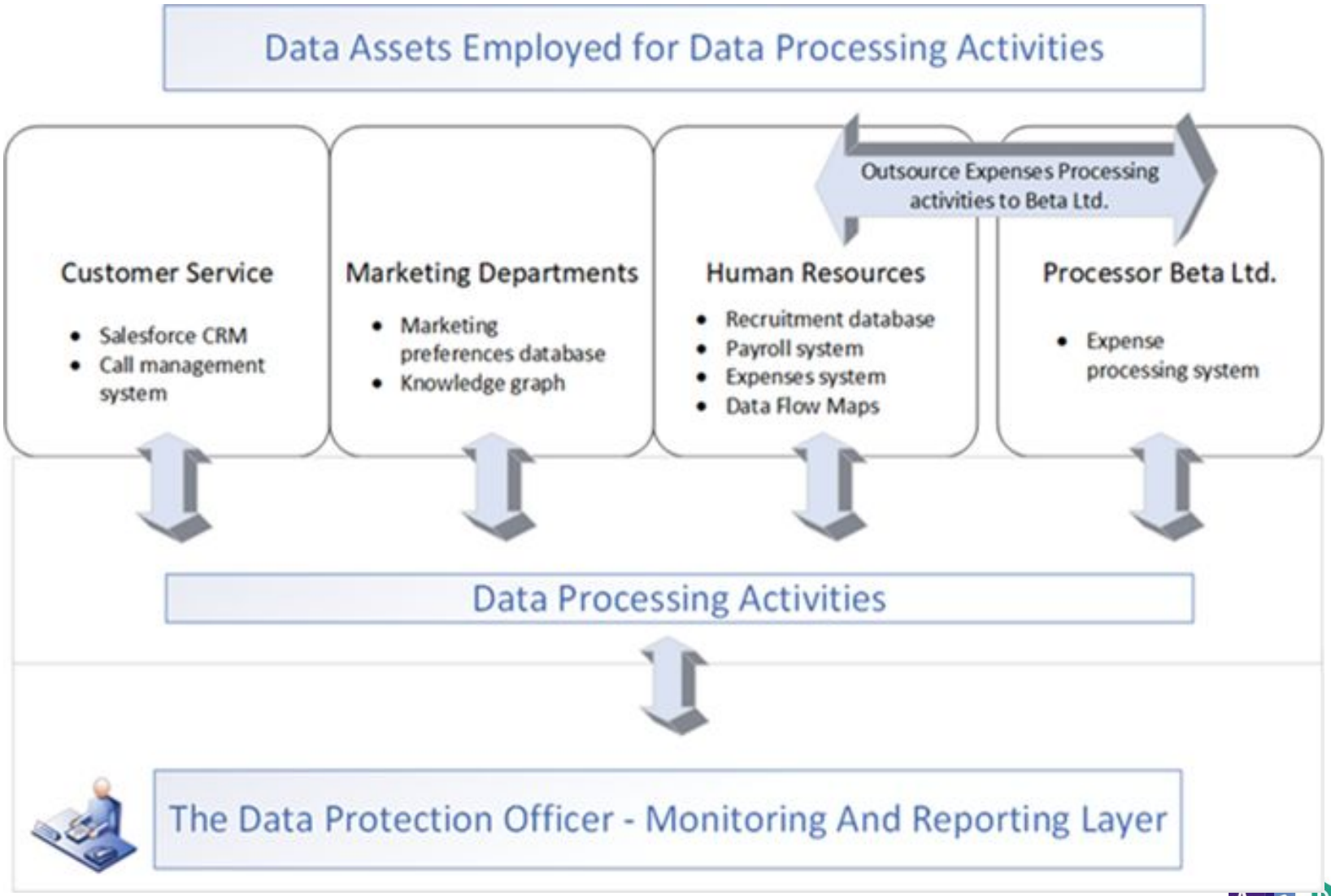
Note: The Property dpv:hasSafeguard and the property range dpv:Safeguard have been submitted to the Data Privacy Community Controls Group for inclusion in the DPV vocabulary.

DPCat specification for ROPA datasets



"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Use case – Alpha Ltd.



"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Use Case - Alpha Ltd.

- Each department maintains its records in a separate catalog
- The organisation's catalog references these as datasets.
- This information maintained in a department's catalog and records fields are produced based on how they conduct their activities.
- The outcome is an RDF graph used in the catalog records maintained in GraphDB
- SPARQL queries were then used to create 'views' for the generation of a ROPA

The catalog, datasets, queries, and outputs for this use case are available here: https://github.com/coolharsh55/DPCat.

Sample Extract of Controller ROPA - Alpha Ltd.

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

https://github.com/coolharsh55/DPCat.

"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Outcome of Use case

Requirement	How DPCat met the expectations
Supports the heterogeneity of data sources	Achieved lightweight integration of diverse data on data processing activities and easy interoperability due to DCAT standard
Enable standards-based collation of the data required for completion of a ROPA	Metadata-level integration sufficient for basic ROPA functions , and reduces need for detailed data alignment
Record temporal validity of processing activities	DPCat provides start date and end date of processing activities. Any new data processing can is easily identifiable.
Support periodic or continuous changes to data processing activity	
Record identity of activity host and organisational unit and relevant contact	DPCat provides publisher and contact name
Facilitate searching records, e.g. identify activities active on a specific date	Power full-text search in catalogs available with CKAN
Enable the creation of ROPA and other compliance-related documentation using information collected in the records	ROPA successfully generated
Minimise the data to be collected and integrated	Can reuse data catalog implementations for easy/low-cost deployment
Easy to deploy, e.g. based on established or commonly used software platforms	

Conclusions

- Our research sought to establish the extent to which implementing a Data Processing Activities catalog based on DCAT-AP and DPV can overcome the heterogeneity of sources to facilitate the preparation of a ROPA. For this, we presented a use case and developed a prototype system to catalog the organisation's diverse data processing activities using SPARQL queries to output a ROPA document.
- A first step towards handling the heterogeneity of data sources representing the organisation's data processing activities presents significant challenges when completing a ROPA.
- DPCat provides a lightweight, low cost, and metadata-level integration for compliance information regarding processing activities from heterogeneous sources.
- DPCat solution advances alignments between disciplinary and domain-specific metadata standards.
- DPCat enables data catalog implementations by providing a common interoperable base for ROPA without requiring full alignment or merging all the underlying data sources.

"Building a Data Processing Activities Catalog: Representing Heterogeneous Compliance-related Information for GDPR using DCAT-AP and DPV"
Paul Ryan, Harshvardhan J. Pandit, Rob Brennan | Paul.Ryan@dcu.ie | SEMANTiCS2021 | Tuesday SEP-07 2021 | https://doi.org/10.3233/SSW210043

Acknowledgements

Questions ?