



Engaging Content
Engaging People

Workshop on Ontology Design Patterns 2018
Co-located with ISWC 2018, Monterey, California, USA

C.C. - BY - 4.0

An Ontology Design Pattern for Describing Personal Data in Privacy Policies

Harshvardhan J. Pandit, Declan O' Sullivan, Dave Lewis

[@coolharsh55](https://twitter.com/coolharsh55) pandith@tcd.ie → email

ADAPT Centre - Trinity College Dublin - Ireland

<https://openscience.adaptcentre.ie/> → check out
our work



A Legal Document that
“may” or “may not”
inform you about the
privacy of your data

- monolithic text document
- verbose | long length
- difficult to read & understand



“Can’t say. It’s private.”

“An ODP for Describing Personal Data in Privacy Policies”

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

1. Usable Privacy Project → uses semantic web!
<https://usableprivacy.org/>
2. Terms of Service; Didn't Read → crowdsourced
<https://tosdr.org>
3. Privacy Guide → dashboard, similar to #1
Tsfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J.: *I Read but Don'T Agree: Privacy Policy Benchmarking Using Machine Learning and the EU GDPR*. In: Companion Proceedings of the The Web Conference 2018. pp. 163–166.
4. Pribot → state of the art
<https://pribot.org/>

"An ODP for Describing Personal Data in Privacy Policies"

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

- different approaches → same information
- same information → different 'ontologies'
- different 'ontologies' → same underlying concepts

models of representation

*context
i.e. representing
information within
privacy policy*

"An ODP for Describing Personal Data in Privacy Policies"

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

1. Information about Personal Data
2. Information within Privacy Policy
3. From GDPR-ready policies

The screenshot shows the 'Privacy Policy' page for Airbnb Ireland. It includes a legend on the right side with the following categories: data category, data type, process, automated, legal basis, data source, data retention, processor, third-party, data-sharing, consent, rights, and location. The document text is annotated with colored boxes and labels corresponding to these categories. For example, 'three' is labeled as 'data type', 'necessary' as 'data type', 'Account Information' as 'data category', and 'financial information' as 'data type'. The text describes the collection and use of personal information, including account details, profile information, identity verification, and payment information.

Airbnb Ireland Privacy Policy

“An ODP for Describing Personal Data in Privacy Policies”

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh5](https://twitter.com/coolharsh5)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

1. What personal data is **collected**? e.g. email
2. Does the data have a **category**? e.g. contact information
3. What was its **source**? e.g. user
4. How is it **collected**? e.g. given by user, automated
5. What is it **used** for? e.g. creating an account, authentication and verification
6. How long is it **retained** for? e.g. 90days after account deletion
7. Who is it **shared** with? e.g. name of partner organisation(s)
8. What is the **legal basis**? e.g. given consent, legitimate use
9. What processes/**purposes** was the data shared for? e.g. analytics, marketing
10. What is the legal type of **third party**? e.g. processor, controller, authority
11. How can personal data be rectified or corrected?
12. How can personal data be deleted or removed?
13. How can a copy of the personal data be obtained?
14. How can personal data be transferred to another party?
15. How can information about the personal data be obtained?

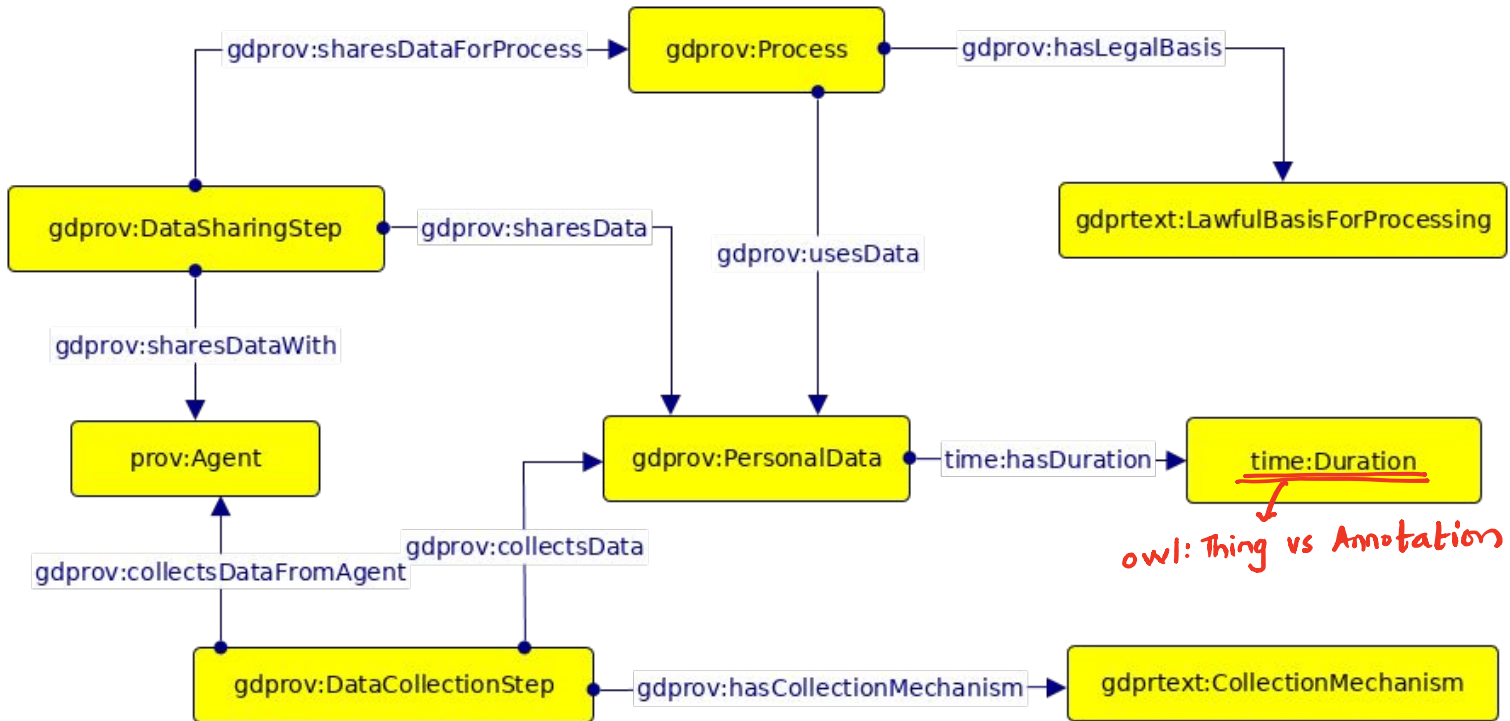
additional questions common over privacy policy

"An ODP for Describing Personal Data in Privacy Policies"

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit





owl: Thing vs Annotation

Reuses vocabularies

GDPROV - GDPR Provenance Vocabulary

GDPRTEXT - GDPR text extensions

PROV-O - Provenance vocabulary

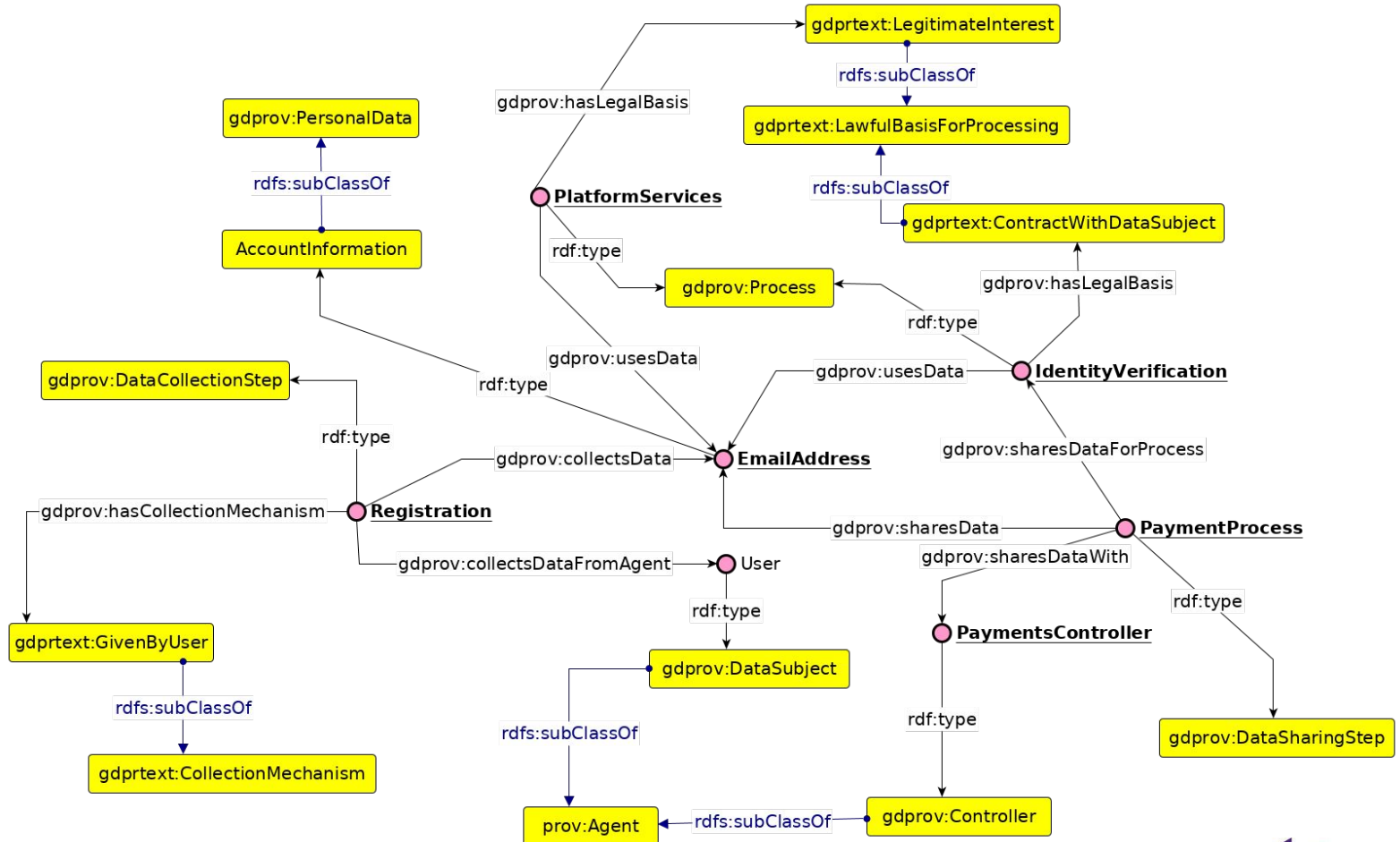
"An ODP for Describing Personal Data in Privacy Policies"

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

#7





"An ODP for Describing Personal Data in Privacy Policies"

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

1. Level of abstraction; e.g. storage, sharing, time → Time Vocabulary?
 2. Privacy Policy as a document ↙ model document
↘ model contents
↘ model policy
 3. Break into Smaller Patterns???
- ↓
- more verbase
 - modular
 - difficult to integrate

Problem Domain

- Privacy Policy is a legal document that specifies information regarding the use of Personal Data amongst other things.
- This information needs to be represented as machine-readable metadata to enable machine-based approaches over it.
- Approaches that work with Privacy Policies have their own internal representation of this metadata specific to their aims.
- It is difficult to garner consistency and standardisation on privacy policies as a document itself.

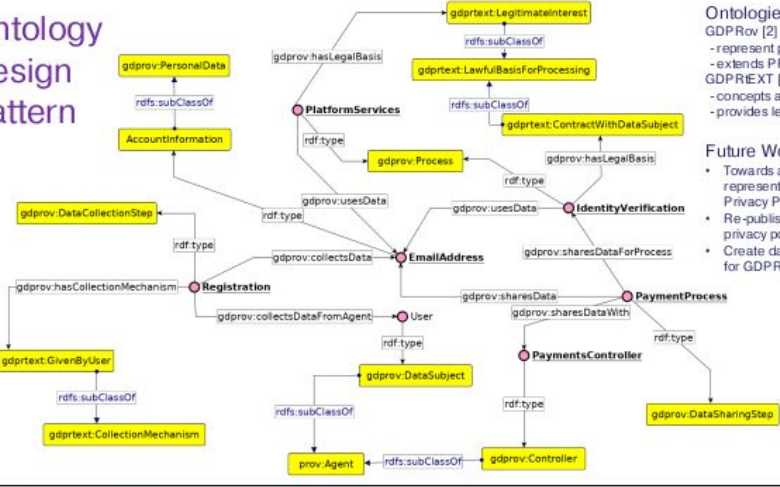


Solution Sphere

1. Identify relevant information that is common to all privacy policies.
2. Incorporate legal requirements from General Data Protection Regulation (GDPR).
3. Use existing vocabularies to represent concepts and relations.
4. Define an ODP that can be incorporated into ontologies or used as a mapping between approaches.
5. Create a corpus of privacy policy metadata using ODP → future work!

--- MORE AT POSTER SESSION ---

Ontology Design Pattern



Ontologies Re-used

- GDPRov [2]
- represent provenance
- extends PROV-O [3]; P-Plan [4]
- GDPREXT [5]
- concepts and terms from GDPR
- provides legal basis

Future Work

- Towards an ontology for representing the entirety of Privacy Policy as a document.
- Re-publish existing corpus of privacy policies using ODP.
- Create dataset from new policies for GDPR (post 25 May 2018)

“An ODP for Describing Personal Data in Privacy Policies”

<http://openscience.adaptcentre.ie/> | pandith@tcd.ie | [@coolharsh55](https://twitter.com/coolharsh55)

WOP 2018 (ISWC 2018) ; Monterey California, USA ; Presented by: Harshvardhan J. Pandit

#10

