A New Generic Representation for Modeling Privacy

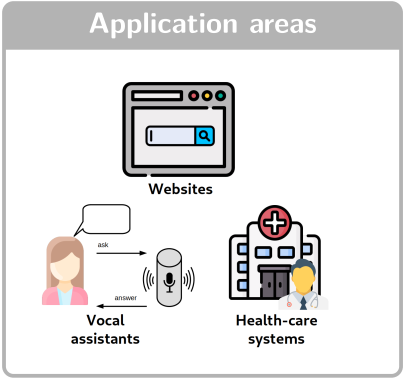
Myriam Clouet,
Thibaud Antignac, Mathilde Arnaud, Gabriel Pedroza, Julien Signoles

Université Paris-Saclay, CEA, List, France

firstname.lastname@cea.fr

Context

Privacy in information systems



Application areas

Websites

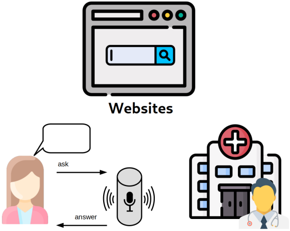
Vocal assistants

Health-care systems

⚠ **Personal data** processing

Privacy in information systems

Application areas



Websites

Vocal assistants

Health-care systems

Laws & Regulations

- GDPR
- Australian Privacy Act
- Japanese Privacy Act
- ...

⚠ **Personal data** processing

⚠ **Mandatory!**

Judgments for non-compliance with the GDPR

Google

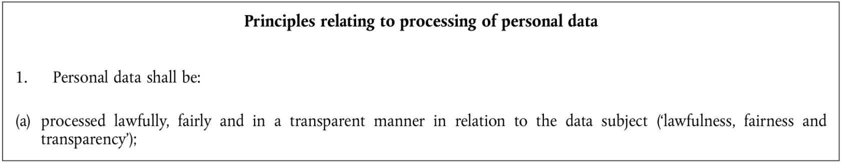
💵 **50 million €** (2019) [15]

WhatsApp

💵 **225 million €** (2021) [14]

Fines for **invalid consent** (lack of transparency)

GDPR



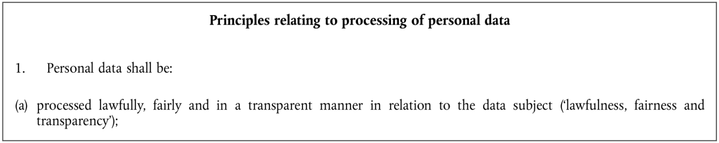
Principles relating to processing of personal data

1. Personal data shall be:

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

Privacy verification - Complex

Document



Principles relating to processing of personal data

1. Personal data shall be:

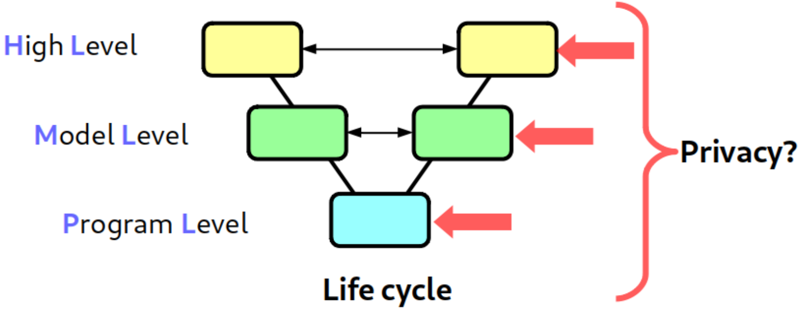
(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

Gap

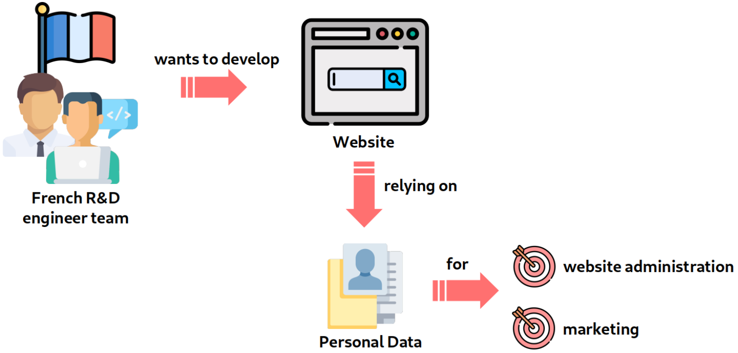
Code

```
public Combined[] join(Payroll[] Ps, Employee[] Es) {
   Combined tab[] = new Combined[Ps.length];
   for (int i=0; i < Ps.length; i++)
      if (Ps[i] != null) tab[i] = checkJoinIndAndfindEmployee(Ps[i], Es);
      else tab[i] = null;
   return tab;}
```

Privacy verification - Complex



Running example



Example: users's e-mail addresses

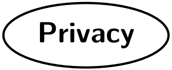
- inform subscription expiration
- send targeted advertising

Contributions

Contributions

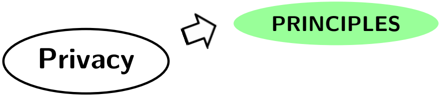
New Generic Privacy Representation

Existing representations



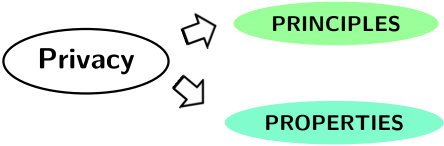
Privacy

Existing representations



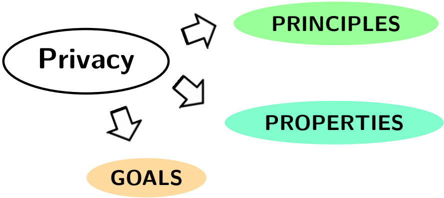
🟩 Principles [6, 26, 13]

Existing representations



🟢 Principles [6, 26, 13]
🟢 Properties [18]

Existing representations

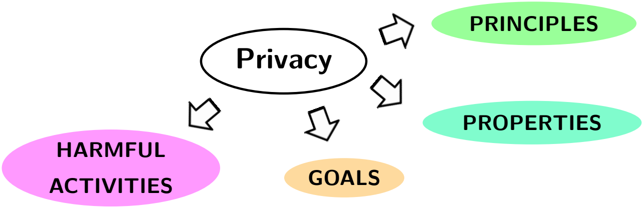


● Principles [6, 26, 13]

● Properties [18]

● Goals [1]

Existing representations


Privacy

PRINCIPLES

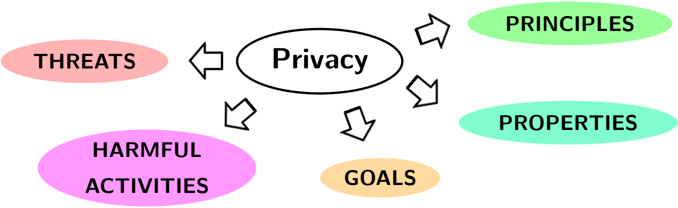
PROPERTIES

GOALS

HARMFUL ACTIVITIES

🟩	Principles [6, 26, 13]	🟪	Harmful Activities [21]
🟦	Properties [18]		
🟧	Goals [1]		

Existing representations



Privacy

THREATS

PRINCIPLES

PROPERTIES

GOALS

HARMFUL ACTIVITIES

🟩 Principles [6, 26, 13]

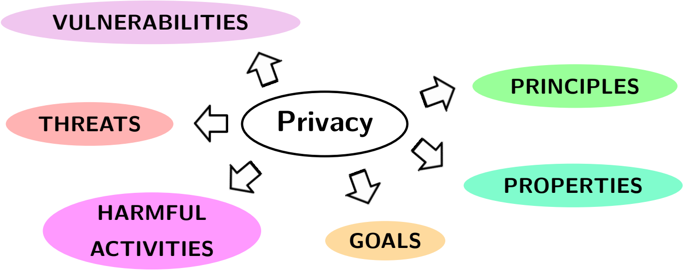
🟦 Properties [18]

🟧 Goals [1]

🟪 Harmful Activities [21]

🟥 Threats [8]

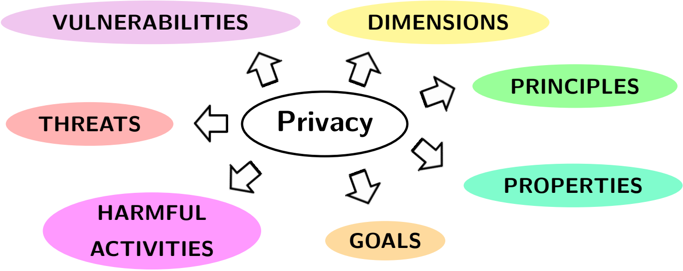
Existing representations



● Principles [6, 26, 13]
● Properties [18]
● Goals [1]

● Harmful Activities [21]
● Threats [8]
● Vulnerabilities [1]

Existing representations



- ● Principles [6, 26, 13]
- ● Properties [18]
- ● Goals [1]
- ● Harmful Activities [21]
- ● Threats [8]
- ● Vulnerabilities [1]
- ● Dimensions [4]

Existing representations observations

- Many **specific** representations

- Many rely on **similar notions**
  example:
  - Principle: *Purpose limitation* [6]
  - Property: *Purpose binding* [10]
  - Goal: *Choice/Consent* [1]
  - Harmful Activity: *Secondary Use* [21]

Positioning

	Existing representations
Adapted to specific situations	✓
Genericity	✗
Comparing papers	✗
Identifying key elements	?

Positioning

	Existing representations
Adapted to specific situations	✓
Genericity	✗
Comparing papers	✗
Identifying key elements	?

💡 **Solution:** A new **generic** representation

GePyR -
A Generic Privacy Representation

Group via generic **categories**

GePyR - A Generic Privacy Representation

Group via generic **categories**

- **Confidentiality category**

the visibility of the e-mail addresses

GePyR -
A Generic Privacy Representation

Group via generic **categories**

- Confidentiality category
- Consent category

  agreement between the website owner and its users

GePyR - A Generic Privacy Representation

Group via generic **categories**

- **Confidentiality category**
- **Consent category**
- **Transparency category**

  awareness of the users

GePyR - A Generic Privacy Representation

Group via generic **categories**

- **Confidentiality category**
- **Consent category**
- **Transparency category**
- **Accountability category**

  ability for the website owner to demonstrate data processing rule respect

Specialisations - Consent

Principles	Properties	Goals	Harmful Activities	Threats	Vulnerabilities	Dimensions
Purpose Limitation [6, 26, 13]	Purpose Binding [10]	Choice/Consent [1]	Interrogation [21]	Policy and Consent Non-Compliance [8]	Information Collection [1]	Purpose [4]
Storage Limitation [6]	Necessity of Data Collection and Processing [10]		Secondary Use [21]		Solicitation [1]	Retention [4]
Data Minimization [6, 13]					Information Monitoring [1]	
Accuracy [6]					Information Storage [1]	

Specialisations - Consent

Principles	Properties	Goals	Harmful Activities	Threats	Vulnerabilities	Dimensions
Purpose Limitation [6, 26, 13]	Purpose Binding [10]	Choice/Consent [1]	Interrogation [21]	Policy and Consent Non-Compliance [8]	Information Collection [1]	Purpose [4]
Storage Limitation [6]	Necessity of Data Collection and Processing [10]		Secondary Use [21]		Solicitation [1]	Retention [4]
Data Minimization [6, 13]					Information Monitoring [1]	
Accuracy [6]					Information Storage [1]	

Classification example - Consent

LVL	TARGET	REPRESENTATION		REF
HL	Mobile App	Goals	Choice/Consent	[19]
	Home automation	Principles	Purpose limitation	[7]
	Web sites	Principles	Lawfulness, fairness and transparency	[15]
ML	Hospital Information System	Harmful Activities	Secondary use	[17]
	Diagnostic process	Threats	Policy and consent non-compliance	[25]
	Smart device (IOT)	Principles	Lawfulness, fairness and transparency	[3]
PL	Hospital Information System	Threats	Policy and consent non-compliance	[24]
	Web sites	Principles	Purpose limitation	[11]
	Database	Harmful Activities	Secondary use	[9]

Classification example - Consent

LVL	TARGET	REPRESENTATION		REF
HL	Mobile App	Goals	Choice/Consent	[19]
	Home automation	Principles	Purpose limitation	[7]
	Web sites	Principles	Lawfulness, fairness and transparency	[15]
ML	Hospital Information System	Harmful Activities	Secondary use	[17]
	Diagnostic process	Threats	Policy and consent non-compliance	[25]
	Smart device (IOT)	Principles	Lawfulness, fairness and transparency	[3]
PL	Hospital Information System	Threats	Policy and consent non-compliance	[24]
	Web sites	Principles	Purpose limitation	[11]
	Database	Harmful Activities	Secondary use	[9]

Classification example - Consent

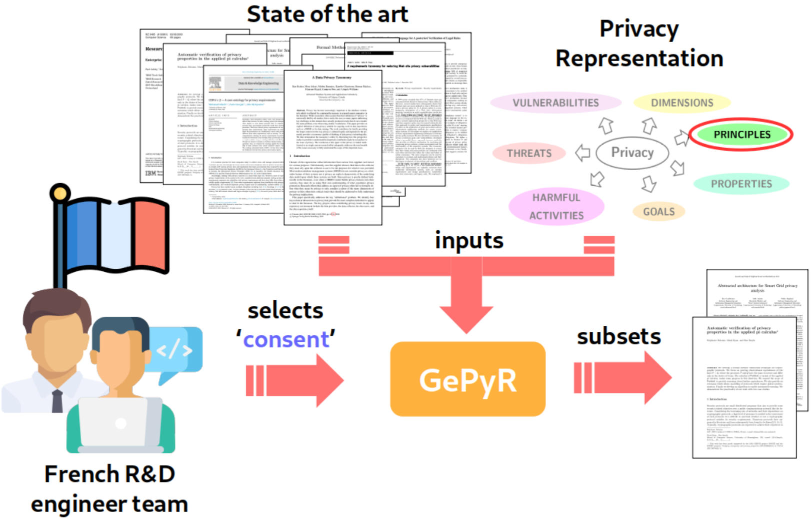
LVL	TARGET	REPRESENTATION		REF
HL	Mobile App	Goals	Choice/Consent	[19]
	Home automation	Principles	Purpose limitation	[7]
	Web sites	Principles	Lawfulness, fairness and transparency	[15]
ML	Hospital Information System	Harmful Activities	Secondary use	[17]
	Diagnostic process	Threats	Policy and consent non-compliance	[25]
	Smart device (IOT)	Principles	Lawfulness, fairness and transparency	[3]
PL	Hospital Information System	Threats	Policy and consent non-compliance	[24]
	Web sites	Principles	Purpose limitation	[11]
	Database	Harmful Activities	Secondary use	[9]

[3] blockchain smart contracts

[11] **J**ava **I**nformation **F**low (JIF)

...

Running example



State of the art

Privacy Representation

VULNERABILITIES

DIMENSIONS

THREATS

Privacy

PRINCIPLES

PROPERTIES

HARMFUL ACTIVITIES

GOALS

inputs

selects 'consent'

GePyR

subsets

Abstracted architecture for Smart Grid privacy analysis

Automatic verification of privacy properties in the applied pi calculus'

French R&D engineer team

Positioning

	Existing representations	GePyR
Adapted to specific situations	✓	✓
Genericity	✗	✓
Comparing papers	✗	✓
Identifying key elements	?	?

Positioning

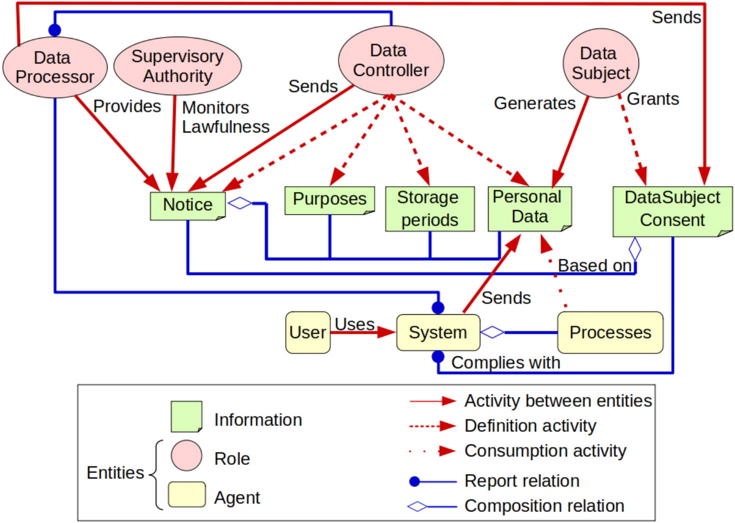
	Existing representations	GePyR
Adapted to specific situations	✓	✓
Genericity	✗	✓
Comparing papers	✗	✓
Identifying key elements	?	?

complete with an **ontology**

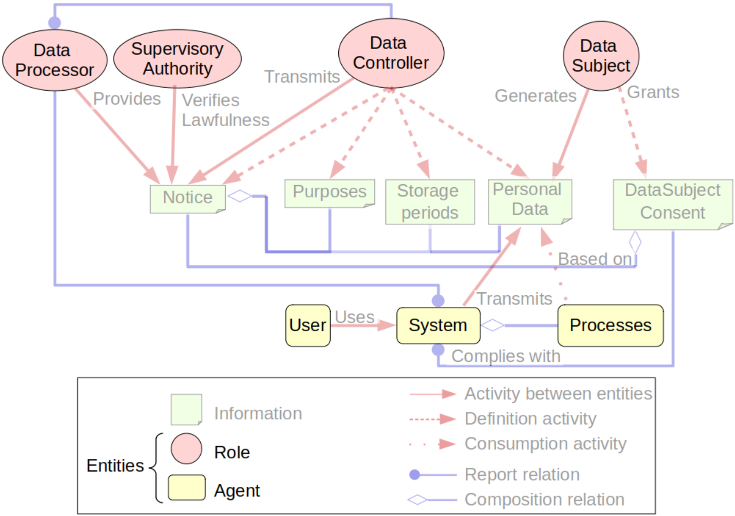
Contributions

New Privacy Context Ontology

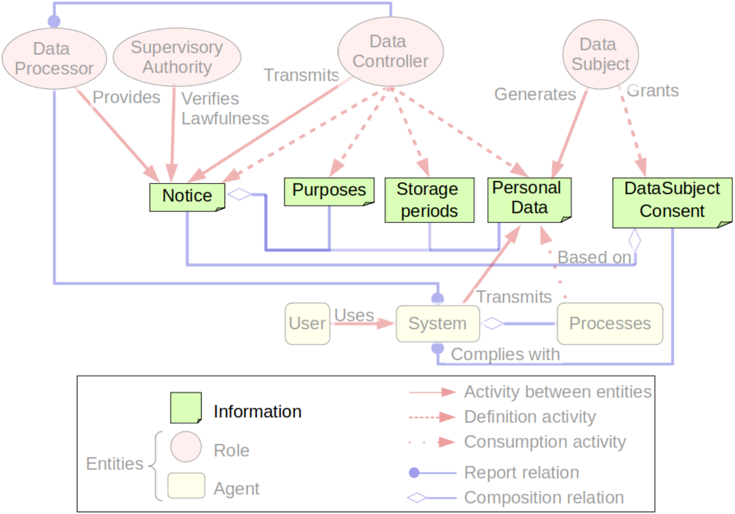
PyCO - Privacy Context Ontology



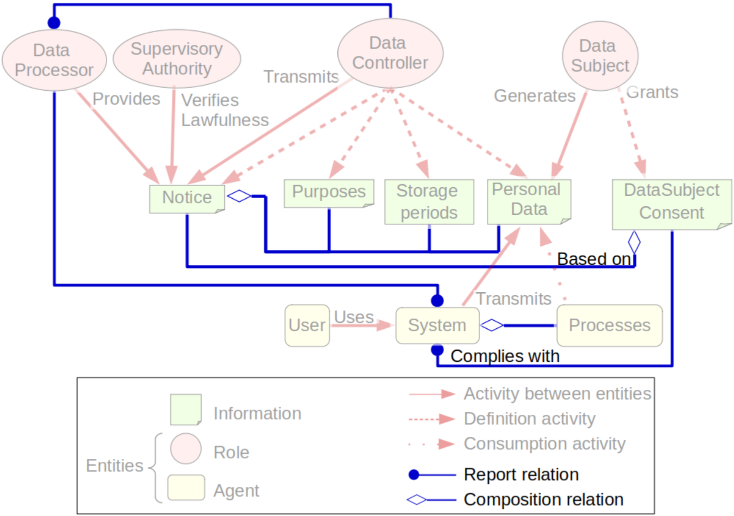
PyCO - Privacy Context Ontology



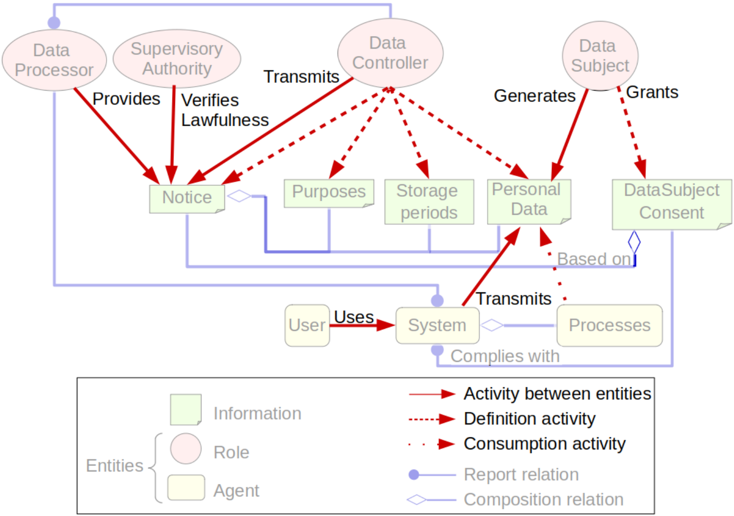
PyCO - Privacy Context Ontology



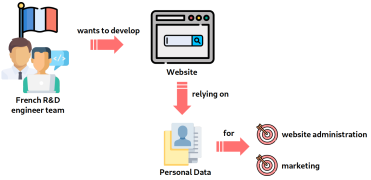
PyCO - Privacy Context Ontology



PyCO - Privacy Context Ontology



PyCO - Running Example



Data Subject User	website's user
Data Controller Data Processor	website's owner
Personal Data	email address
Purposes	administration & marketing
System	website
Processes	functions (website's code)
Supervisory Authority	CNIL

PyCO - Example of use

LVL	DOMAIN	SYSTEM	PRO-CESSES	PUR-POSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

PyCO - Running Example

LVL	DOMAIN	SYSTEM	PRO-CESSES	PUR-POSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

Positioning

	Existing representations	GePyR	PyCO
Adapted to specific situations	✓	✓	✓
Genericity	✗	✓	≈
Comparing papers	✗	✓	✓
Identifying key elements	?	?	✓

Conclusion and Future Work

Conclusion and future work

Contributions *Paper classification*

- **GePyR**: **Ge**neric **P**rivac**y** **R**epresentation
  → *Genericity*
- **PyCO**: **P**rivac**y** **C**ontext **O**ntology
  → *Key element identification*

Future Work

- **Extending** our state of the art
- **Identifying** properties related to our category of consent
- **Defining** a language to verify consent properties

Bibliography

References I

[1] Annie I Anton and Julia B Earp. “A requirements taxonomy for reducing web site privacy vulnerabilities”. In: *Requirements engineering* (2004).

[2] Philip Asuquo et al. “Security and privacy in location-based services for vehicular and mobile communications: an overview, challenges, and countermeasures”. In: *IEEE Internet of Things Journal* (2018).

[3] Masoud Barati et al. “GDPR Compliance Verification in Internet of Things”. In: *IEEE Access* (2020).

[4] Ken Barker et al. “A data privacy taxonomy”. In: *British National Conference on Databases*. 2009.

[5] Michael R Clarkson and Fred B Schneider. “Hyperproperties”. In: *Journal of Computer Security* (2010).

References II

[6] Commission Européenne. *Regulation (EU, General Data Protection Regulation)*. Tech. rep. 2016.

[7] Commission Nationale Informatique et Libertés. *Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*. Tech. rep. 2020.

[8] Mina Deng et al. “A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements”. In: *Requirements Engineering* (2011).

[9] Guillaume Dufay, Amy Felty, and Stan Matwin. “Privacy-sensitive information flow with JML”. In: *International Conference on Automated Deduction*. 2005.

References III

[10] Simone Fischer-Hubner and Amon Ott. “From a formal privacy model to its implementation”. In: *Proceedings of the 21st National Information Systems Security Conference*. 1998.

[11] Katia Hayati and Martin Abadi. “Language-based enforcement of privacy policies”. In: *International Workshop on Privacy Enhancing Technologies*. 2004.

[12] Lucca Hirschi, David Baelde, and Stephanie Delaune. “A method for unbounded verification of privacy-type properties”. In: *Journal of Computer Security* (2019).

[13] Jaap-Henk Hoepman. “Privacy design strategies”. In: *IFIP International Information Security Conference*. 2014.

References IV

[14] Jacques Cheminat. *RGPD : WhatsApp condamné à 225 millions d'euro d'amende*. URL: https://www.lemondeinformatique.fr/actualites/lire-rgpd-whatsapp-condamne-a-225-meteuro-d-amende-84044.html (visited on 09/06/2021).

[15] Nicolas Certes. *RGPD : Google condamné à 50 millions d'euro par la CNIL*. URL: https://www.lemondeinformatique.fr/actualites/lire-rgpd-google-condamne-a-50-meteuro-par-la-cnil-74062.html (visited on 12/10/2020).

[16] Dr Ian Oliver. *Privacy engineering: A dataflow and ontological approach*. CreateSpace Independent Publishing Platform, 2014.

References V

[17] Milan Petkovic, Davide Prandi, and Nicola Zannone. "Purpose control: Did you process the data for the intended purpose?" In: *Workshop on Secure Data Management*. 2011.

[18] Andreas Pfitzmann and Marit Hansen. "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management". In: (2010).

[19] Marco Robol et al. "Consent Verification Under Evolving Privacy Policies". In: *IEEE 27th International Requirements Engineering Conference (RE)*. 2019.

[20] Tanusree Sharma, John C Bambenek, and Masooda Bashir. "Preserving Privacy in Cyber-physical-social Systems: An Anonymity and Access Control Approach". In: (2020).

References VI

[21] Daniel J Solove. “A taxonomy of privacy”. In: (2005).

[22] Degang Sun et al. “Secure HybridApp: A detection method on the risk of privacy leakage in HTML5 hybrid applications based on dynamic taint tracking”. In: *2nd IEEE International Conference on Computer and Communications (ICCC)*. 2016.

[23] The AVISPA team. *HLPSL Tutorial*. Tech. rep. 2006.

[24] Shukun Tokas, Olaf Owe, and Toktam Ramezanifarkhani. “Language-based mechanisms for privacy-by-design”. In: *IFIP International Summer School on Privacy and Identity Management*. 2019.

References VII

[25] Michael Carl Tschantz, Anupam Datta, and Jeannette M Wing. “Formalizing and enforcing purpose restrictions in privacy policies”. In: *2012 IEEE Symposium on Security and Privacy*. 2012.

[26] Working Party 29. *ARTICLE 29 DATA PROTECTION WORKING PARTY - Opinion 03/2013 on purpose limitation*. Tech. rep. 2013.

Crédit - Images

Icones faites par :

- **Freepik**
- **xnimrodx**
- **Vitaly Gorbachev**
- **Pixel perfect**
- **Flat Icons**
- **Vectors Market**
- **pongsakornRed**
- **Eucalyp**
- **Smashicons**

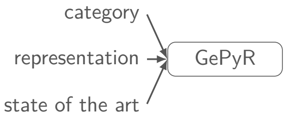
disponibles sur **www.flaticon.com**

Appendix

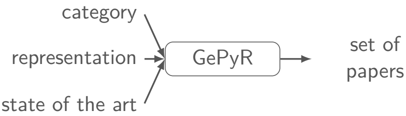
Appendix

GePyR with PyCO

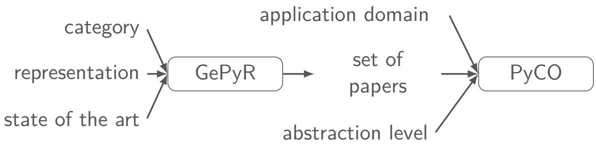
Methodology example - GePyR & PyCO



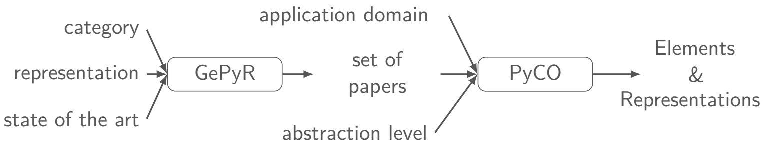
Methodology example - GePyR & PyCO



Methodology example - GePyR & PyCO



Methodology example - GePyR & PyCO



Appendix

Confidentiality category

Specialisations - Confidentiality

Principles	Properties	Goals	Harmful Activities	Threats	Vulnerabilities	Dimensions
Integrity and Confidentiality [6]	Anonymity [18]	Integrity/Security [1]	Identification [21]	Linkability [8]	Information Aggregation [1]	Visibility [4]
	Unlinkability [18]		Breach of Confidentiality [21]	Identifiability [8]		Granularité [4]
Data Breach Notification [13]	Undetectability [18]			Compliance [8]	Information Transfer [1]	
	Unobservability [18]		Disclosure [21]	Detectability [8]		
			Increased Accessibility [21]	Disclosure of Information [8]		

Classification example - Confidentiality

LVL	TARGET	REPRESENTATION		REF
HL	Location-based services	Threats	Linkability	[2]
	Communication protocols	Threats	Disclosure of information	[23]
	Trace sets	Properties	Non-interference	[5]
ML	Communication protocols	Properties	Unlinkability	[12]
	Data-flow diagram	Vulnerabilities	Information Storage	[16]
	Cyber Physical Systems	Threats	Disclosure of information	[20]
PL	Internet of Things	Principles	Data minimization	[3]
	Web privacy policies	Properties	Non-Disclosure	[11]
	Mobile applications	Principles	Data breach notification	[22]

Classification example - Confidentiality

LVL	TARGET	REPRESENTATION		REF
HL	Location-based services	Threats	Linkability	[2]
	Communication protocols	Threats	Disclosure of information	[23]
	Trace sets	Properties	Non-interference	[5]
ML	Communication protocols	Properties	Unlinkability	[12]
	Data-flow diagram	Vulnerabilities	Information Storage	[16]
	Cyber Physical Systems	Threats	Disclosure of information	[20]
PL	Internet of Things	Principles	Data minimization	[3]
	Web privacy policies	Properties	Non-Disclosure	[11]
	Mobile applications	Principles	Data breach notification	[22]

[23] role-based language to specify communication protocols (with tool)

[20] anonymity technique & purpose-based access control algorithms

Appendix

PyCO

PyCO - Example of use

LVL	DOMAIN	SYSTEM	PRO-CESSES	PUR-POSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

PyCO - Example of use

LVL	DOMAIN	SYSTEM	PRO-CESSES	PUR-POSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

PyCO - Example of use

LVL	DOMAIN	SYSTEM	PRO-CESSES	PUR-POSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

PyCO - Example of use

LVL	DOMAIN	SYSTEM	PROCESSES	PURPOSES	PRIVATE DATA	STORAGE PERIODS	REF
HL	Public Transport	Mobile App	DNL*	Keywords	DNL*	∅	[19]
	Home Automation	Vocal Assistants	DNL*	DNL*	DNL*	∅	[7]
	Web Services	Web Sites	DNL*	DNL*	DNL*	∅	[15]
ML	Medical	IT	BPM**	Keywords	Keywords	∅	[17]
	Medical	Diagnostic Process	Markov Decision Process	Decision function	Keywords	∅	[25]
	Smart Building	Smart device (IOT)	BPM**	Keywords	Keywords	∅	[3]
PL	Medical	IT	Functions	Keywords	Keywords	∅	[24]
	Web services	Web sites	Functions	DNL* or Keywords	"Object"	Keywords	[11]
	Human resources	Database	Functions	Keywords	"Objects"	∅	[9]

* Descriptions in Natural Language

** Business Process Models

Commissariat à l'énergie atomique et aux énergies alternatives
CEA Tech List
Centre de Saclay | 91191 Gif-sur-Yvette Cedex
www-list.cea.fr

Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019