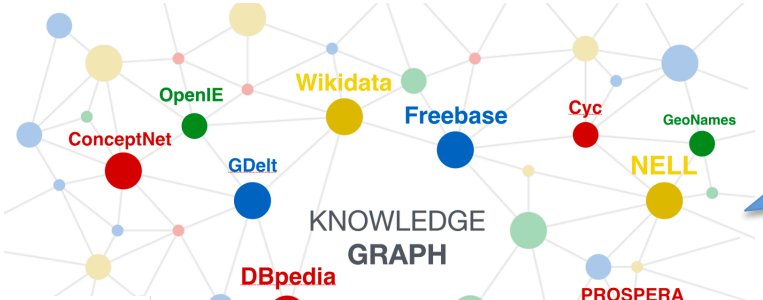
Privacy and Constrained Access

Sabrina Kirrane
Vienna University of Economics and Business

Why would you want to constrain access to a knowledge graph?

Access Control Policies



Terms of use in the form of licenses

Usage Constraints

Enterprise Knowledge Graphs

Efficient Enterprise Data Management reduces costs, increases performance levels and generates additional value for customers and enterprises. However, an obstacle for increasing efficiency in this regard is the disconnected nature of heterogeneous data sources and the exponentially increasing data volumes. As a result, enterprises require information management solutionsthat facilitate the integration and interlinking of heterogeneous data and thus deliver profitable insights and generate Big Data value, conferring long-term competitive advantage.

The Linked Data paradigm provides a semantic coherence layer to existing enterprise systems and data lakes through the introduction of an integrated set of vocabularies, based on which different types of data can be semantically described and to a large extent automatically interlinked, even when retained at source. The realisation of such a Semantic Data Lake for a specific company results in an »Enterprise Knowledge Graph (EKG)«, which can also comprise linked open data, thus extending the scope of the business application areas.

Contact

Dr. Simon Scerri
Senior Postdoctoral Researcher

Fraunhofer Institute for Intelligent Analysis and Information Systems
IAIS
Schloss Birlinghoven
53757 Sankt Augustin, Germany

https://www.iais.fraunhofer.de/en/business-areas/enterprise-information-integration/enterprise-knowledge-graphs.html

snips

Snips

Confused about which technology vendor to use for your business problem? Ask for help

Snips
11-50 Employees

Description:
Making technology disappear. This is what Snips sets to achieve by embedding an Artificial Intelligence (AI) in every connected device. Whether it is a smartphone, a smartwatch, a connected car or a home appliance, they will one day be able to anticipate their owner's intentions, and act preemptively to save time and reduce friction.

Products:
Context Awareness: The user's personal data is turned into a highly contextualized timeline of activity they did during the day. This includes turning location traces into places visited, parsing chat messages to extract people and places, mining emails for hotel and restaurant reservations, and much more! This runs fully on-device, with no user data being sent to our servers, ensuring privacy by design.

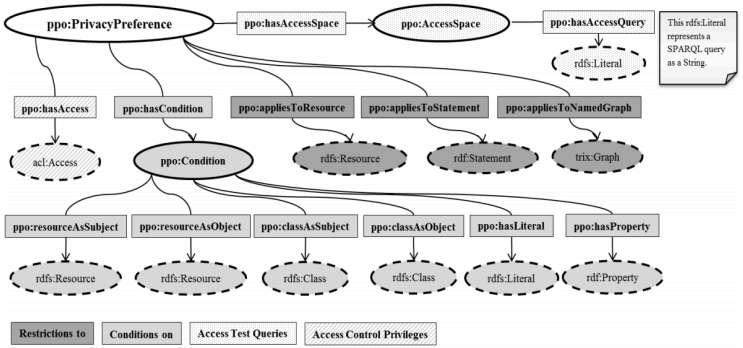
Personal Knowledge Graph: The user's activity timeline and all other contextual data are linked to create his Personal Knowledge Graph.

Headquartered: Batiment C Paris, France
Date Founded: 2013

Contact This Company

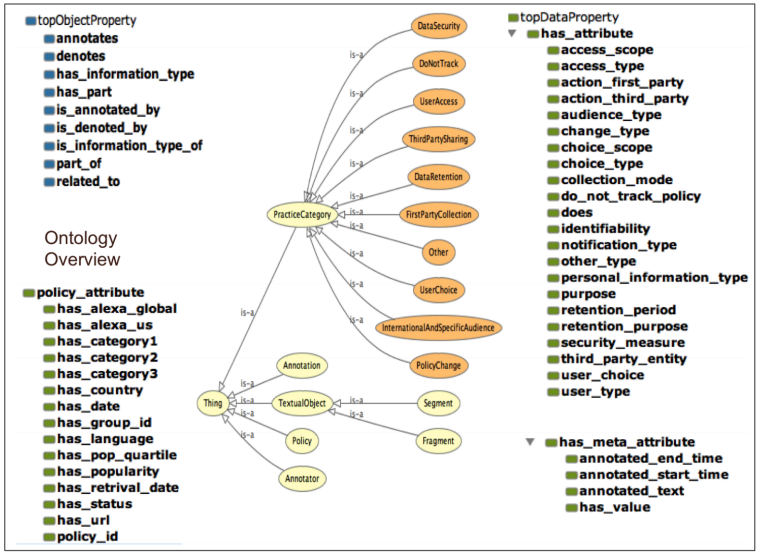
https://www.techemergence.com/companies/snips/

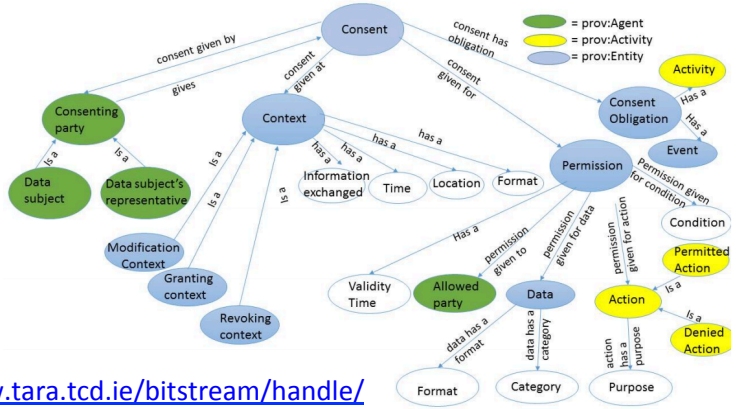
How can I specify constraints?



http://ceur-ws.org/Vol-813/ldow2011-paper01.pdf

There are already several existing ontologies that you could leverage





http://www.tara.tcd.ie/bitstream/handle/2262/82659/88248c5b669175f267069c3319d9ac2d3e84.pdf?sequence=1

http://www.semantic-web-journal.net/system/files/swj1597.pdf

How can I specify constraints?

The SPECIAL Usage Policy Language

version 0.1

Unofficial Draft 06 April 2018

Editor:
Javier D. Fernández (Vienna University of Economics and Business)

Authors:
Piero Bonatti (Università di Napoli Federico II)
Sabrina Kirrane (Vienna University of Economics and Business)
Iliana Mineva Petrova (Università di Napoli Federico II)
Luigi Sauro (Università di Napoli Federico II)
Eva Schlehahn (Unabhängies Landeszentrum für Datenschutz (ULD))

Abstract

This document specifies usage policy language of SPECIAL. The usage policy language is meant to express both the data subjects' consent and the data usage policies of data controllers in formal terms, understandable by a computer, so as to automatically verify that the usage of personal data complies with data subjects' consent.

The ontology defined in this document is publicly available at http://www.specialprivacy.eu/langs/usage-policy.

http://purl.org/specialprivacy/policylanguage

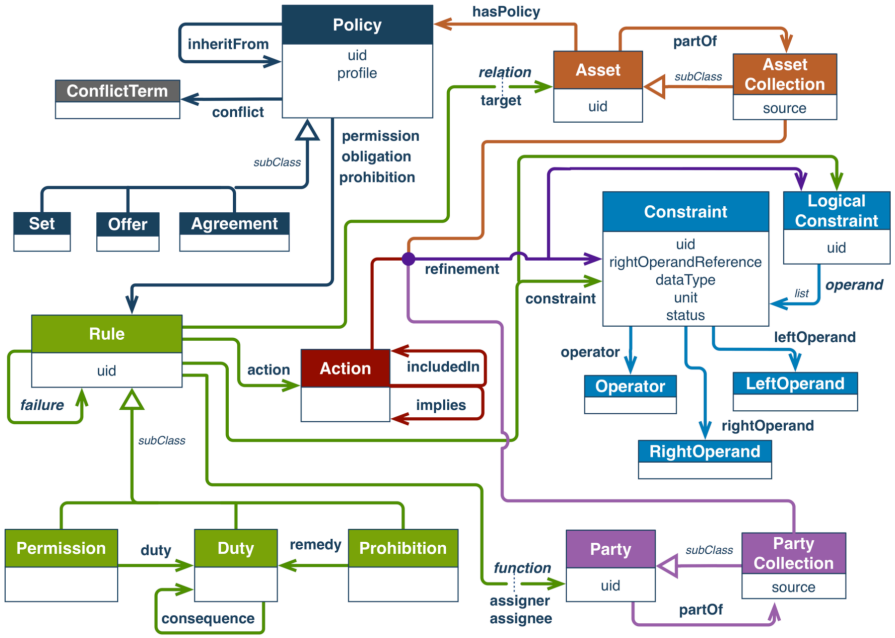
There are a number of policy languages you could choose from!

Table 3
General Policy Languages - Policy Representation and Enforcement

	Policy Type	Policy Representation	Enforcement Mechanism	Enforcement Framework
KAoS [11,12,95]	±authorisations ±obligations	DAML & OWL	DL based enforcement	conflict resolution & harmonisation
Rei [47,48]	±authorisation ±obligations	RDFS, Prolog Rules & OWL	rule based enforcement	dynamic constraints, runtime variables, conflict resolution via metapolicies
Protune [10,7,8]	decision, provisional & abbreviation predicates	lightweight ontologies, rules and meta rules	rule based enforcement	disclosure & negotiation
Proteus [92]	-	policies and domain info as classes, user context as instances	DL & rule based enforcement	conflict resolution & harmonisation, dynamic constraints, runtime variables, disclosure & negotiation
Kolovski et al. [56]	-	XACML policies as DL	DL & rule based enforcement	disclosure, rules for conflict resolution

http://www.semantic-web-journal.net/system/files/swj1280.pdf

How can I specify constraints?



There are also some standard policy languages!

W3C Recommendation

ODRL Information Model 2.2

W3C Recommendation 15 February 2018

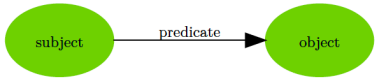
This version:
https://www.w3.org/TR/2018/REC-odrl-model-20180215/

Latest published version:
https://www.w3.org/TR/odrl-model/

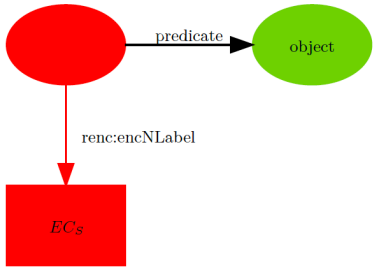
Latest editor's draft:

Expressivity, correctness and completeness with respect to specific use case requirements would need to be investigated!

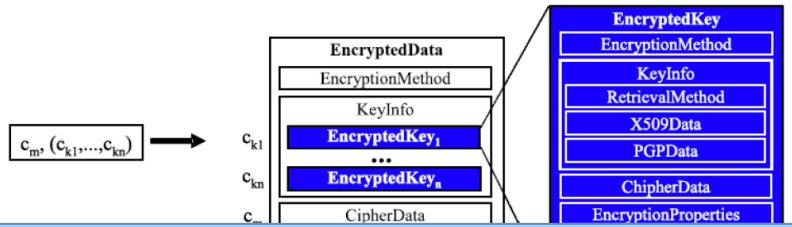
Can’t I simply encrypt the data



(a) Basic Statement



(b) Encryption of the Subject



There are several approaches for encrypting RDF

On Partial Encryption of RDF-Graphs

Mark Giereth

Institute for Intelligent Systems, University of Stuttgart,
70569 Stuttgart, Germany
giereth@iis.uni-stuttgart.de

Abstract. In this paper a method for Partial RDF Encryption (PRE) is proposed in which sensitive data in an RDF-graph is encrypted for a set of recipients while all non-sensitive data remain publicly readable. The result is an RDF-compliant self-describing graph containing encrypted

Efficient querying over encrypted RDF data is still very limited!

How can I handle Regulatory Constraints?

https://plan.io/blog/gdpr-requirements-needed-for-compliance/

It's not just about the GDPR, you may need to consider other legislations also!

Semantic Business Process Regulatory Compliance Checking using LegalRuleML*

Guido Governatori[1], Mustafa Hashmi[1], Ho-Pun Lam[1], Serena Villata[2] and Monica Palmirani[3]

[1] Data61, CSIRO, Spring Hill, QLD 4000, Australia
[2] Université Côte d'Azur, CNRS, Inria, I3S, France
[3] CIRSFID, University of Bologna

Abstract. Legal documents are the source of norms, guidelines, and rules that often feed into different applications. In this perspective, to foster the need of development and deployment of different applications, it is

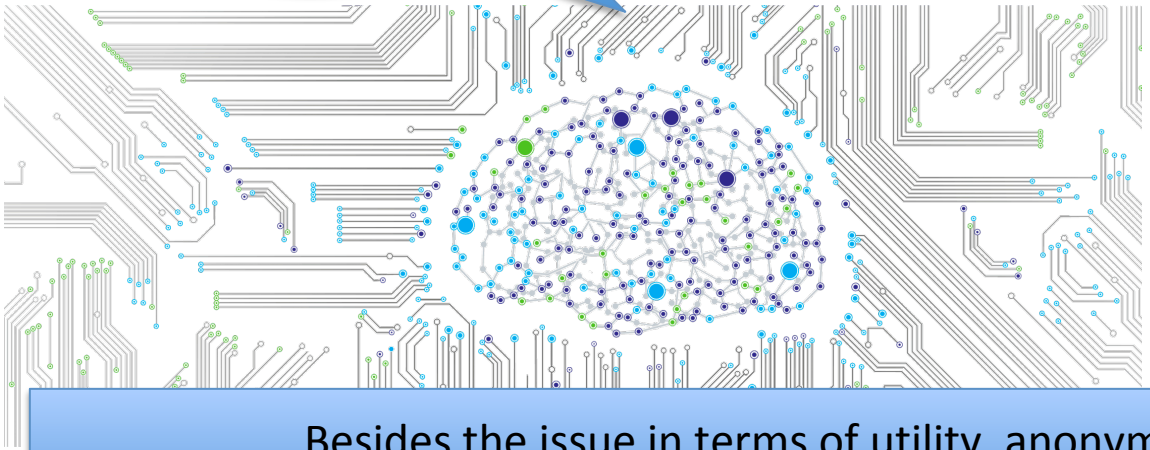
Public Institutions | Banking | Retail & Distribution | Architecture & Engineering | Healthcare | NGOs | Shipping

Automated Compliance checking is still an open research challenge!

Beverage

Can’t I just anonymize the data?

However, anonymisation decreases utility



k — RDF-Neighbourhood Anonymity: Combining Structural and Attribute-Based Anonymisation for Linked Data

Benjamin Heitmann[12], Felix Hermsen[1], and Stefan Decker[12]

[1] Informatik 5 – Information Systems
RWTH Aachen University, Ahornstr. 55, 52056 Aachen, Germany
lastname@dbis.rwth-aachen.de
[2] Fraunhofer Institute for Applied Information Technology FIT
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
firstname.lastname@fit.fraunhofer.de

Abstract. We provide a new way for anonymising a heterogeneous graph containing personal identifiable information. The anonymisation algorithm is called k — RDF-neighbourhood anonymity, because it changes the one hoop neighbourhood of at least k persons inside an RDF graph so that they cannot be distinguished. This enhances the privacy of persons represented in the graph. Our approach allows us to control the loss of information in different parts of the graph to adjust the trade-off between full privacy and data utility. In particular, we can control the weighting of subgraphs induced by individual properties as well as the weighting of attributes represented by literals. To the best of our knowledge, our approach is the first one which considers all subgraphs of an RDF graph at the same time during the anonymisation, instead of projecting the graph into its subgraphs, anonymising each subgraph separately, and them merging the anonymised subgraphs again. In addition, our approach allows partial anonymisation of RDF graphs, for use cases in which only

Besides the issue in terms of utility, anonymisation is not effective in a big data environment!

htt

How can I trust the data in the knowledge graph?

There are some good starting points, however we have new challenges.

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Web Semantics: Science, Services and Agents on the World Wide Web 5 (2007) 58–71

JOURNAL OF Web Semantics
www.elsevier.com/locate/websem

A survey of trust in computer science and the Semantic Web

Donovan Artz, Yolanda Gil *

Information Sciences Institute, University of Southern California, 4677 Admiralty Way, Marina del Rey, CA 90292, United States

Received 9 February 2006; accepted 23 March 2007
Available online 31 March 2007

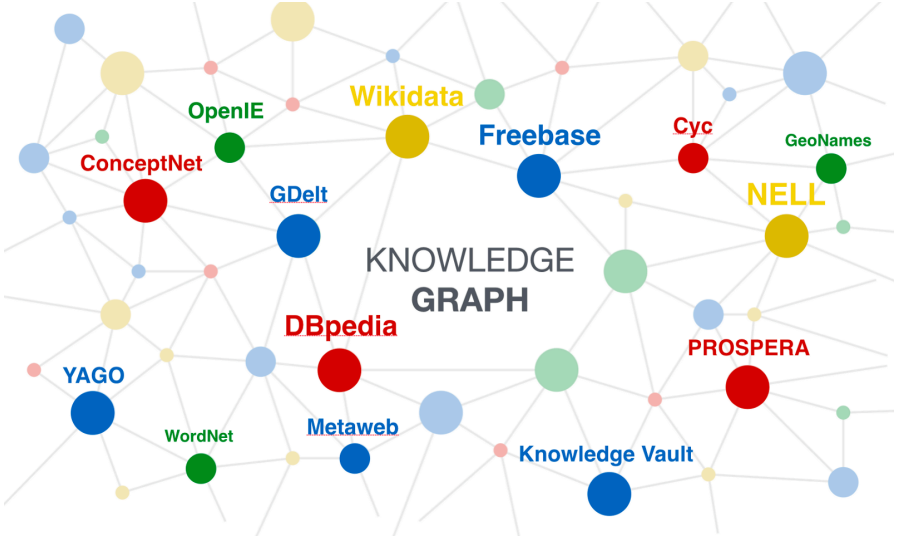
Abstract

Trust is an integral component in many kinds of human interaction, allowing people to act under uncertainty and with the risk of negative consequences. For example, exchanging money for a service, giving access to your property, and choosing between conflicting sources of information all may utilize some form of trust. In computer science, trust is a widely used term whose definition differs among researchers and application areas. Trust is an essential component of the vision for the Semantic Web, where both new problems and new applications of trust are being studied. This paper gives an overview of existing trust research in computer science and the Semantic Web.

Trust mechanisms can be used to validate claims and improve data quality, however we also need to deal with media manipulation (e.g., fake news)

htt
A%20survey%20of%20trust%20in%20computer%20science%20and%20the%20Semantic%20Web.pdf

Key Takeaway



Constraints are a fact of life.

Therefore we need to figure out how to deal with them!