Received: November 25, Date of acceptance: December 15, 2020, Date of publication: December 23 2020, DOI: https://doi.org/10.26826/law-in-context.v37i1.126

Knowledge Graphs: Trust, Privacy, and Transparency from a Legal Governance Approach

By **Daniel Schwabe**[1], Professor at Department of Informatics (INF), Pontifical Catholic University, Rio de Janeiro, **Orcid:** *https://orcid.org/0000-0003-4347-2940*

Carlos Laufer[1], Associate Researcher, TecWeb Lab, at Dept. of Informatics, PUC-Rio, **Orcid:** *https://orcid.org/0000-0002-2606-4067*

and **Pompeu Casanovas**[2], Research Professor at La Trobe University Law School, Bundoora, Melbourne, Australia, **Orcid:** *https://orcid.org/0000-0002-0980-2371*

[1] Pontifical Catholic University, Rio de Janeiro, Brazil

[2] La Trobe University, Melbourne, Australia

ABSTRACT

This paper presents the Knowledge Graph Usage framework, which allows the introduction of Knowledge Graph features to support Trust, Privacy, Transparency and Accountability concerns regarding the use of its contents by applications. A real-world example is presented and used to illustrate how the framework can be used. This article also shows how knowledge graphs can be linked to the elements of legal governance. Thus, it is an invitation to dialogue for legal and Law & Society scholars who might be interested in how the evolution of the web of data and computational sciences intersects with their own discipline.

Keywords – *Knowledge Graphs, Privacy, Transparency, Trust, Legal Governance, Legal interpretation*

Acknowledgements: *A former version of this paper deprived from its present socio-legal dimension has been presented by Daniel Schwabe and Carlos Laufer with the title "Trust and Privacy in Knowledge Graphs" at The Web Conference (WWW'19), May 13, San Francisco, USA. See: Proceedings of WWW '19. ACM, New York, NY, USA. Daniel Schwabe was partially supported by a grant from CNPq. The present work by Pompeu Casanovas has been carried out for the EU H2020 Programme LYNX, Legal Knowledge Graph for Multilingual Compliance Services. The interested reader can find a description of this project as a Research Note at the end of this volume, LiC 37 (1).*

Disclosure statement – *No potential conflict of interest was reported by the authors.*

Suggested citation: Schwabe, D., Laufer, C., Casanovas, P. (2020). "Knowledge Graphs: Trust, Privacy, and Transparency from a Legal Governance Approach", *Law in Context, 37 (1): 24- 41, DOI: https://doi.org/10.26826/law-in-context.v37i1.126*

Summary

1. *Introduction*
2. *A Legal Governance Approach*
   - 2.1 *Components of Legal Governance*
   - 2.2 *A Preliminary Framework*
3. *Background Concepts*
   - 3.1 *Trust*
   - 3.2 *Privacy*
   - 3.3 *Transparency*
4. *An illustration—Disaster Relief Donation*
5. *A Summary of the KG Usage Framework*
   - 5.1 *KI Representation*
   - 5.2 *Controlling Usage*
   - 5.3 *Rules*
   - 5.4 *Conflict Resolution*
6. *Example Scenario revisited*
7. *Conclusions*
8. *References*

1. INTRODUCTION

The advent of the Web has enabled the generation of a vast amount of information, mostly in the form of interlinked pages with texts containing links (references) to other texts. Users accessed information by loading a page - identified by a URL - in their browsers following links in the page to find related information. The advent of social networks added end-user-generated content, further increasing the amount of textual information available.

However, it quickly became evident that search functionality was needed to help users find relevant information, ushering the creation and growth of search engines. While search engines became hugely successful, it also became clear that searching strings contained within Web pages was a limited approach to solve the actual problems that users have, which is to find information that is relevant for a certain need. As stated by Google, people are interested in "things, not strings".[1] For example, users would like to distinguish between "Sydney" as a city and "Sydney" as the name of a known person (e.g., "Sydney Pollack", "Sydney Poitier", "Sydney Lumet"), a football club ("Sydney F..C"), etc…, among many other possible meanings.

Although to this day there isn't a precise definition of the term (Ehrlinger and Wöß 2016, Hogan et al, 2020), we adopt the view that a Knowledge Graph (KG) represents a network of interlinked descriptions of entities (objects, events, concepts etc.)-- a graph-theoretic representation of human knowledge such that it can be ingested with semantics by a machine (Kejriwal 2019).

Graph-based databases have been available for a long time[2] and many more continue to be created on a regular basis – see (Hogan et al, 2020) for a survey. The original vision for the WWW was later enriched to form the Semantic Web, instances of which can also be regarded as a KG—see, for example, the survey by Gandon (2018).

While the graph model or some variant has been used in several KGs, it has already been observed that using only atomic (indivisible) nodes as the "granule" of information is insufficient to express complex types of information, such as events, or time-varying data. For example, Wikidata (Vrandečić and Krötzsch 2014) is organized around Items described by a collection statements (Erxleben et al. 2014). Another reason for having more complex "granules" is recording provenance (meta) data, which is a fundamental part of data in some domains such as life-sciences (Kuhn et al. 2018).

KGs differ also on the way they are built and populated. A few are curated (e.g., CYC), others rely on crowdsourced information (e.g. Wikidata, and several, perhaps most, extract information from structured, semi-structured or textual information harvested from the Web).

The multiplicity of sources and various extraction approaches naturally raises the issue of data quality and confronts the user of the data in the KG with the issue of trusting, or not, the information contained in the KG. For some types of information, for example in case of online reviews and online and social media, this trust can have a direct effect on commercial success (e.g. Angella and Johnson 2016). This highlights the fact that data ultimately expresses a belief, opinion or point of view of some agent – the author.

From a broader perspective, information (and knowledge) is said to have become the prime resource in the Third Industrial Revolution, also called the Digital Age – when digital technologies enabled new ways of generating, processing and sharing information (Castells 2010, Rifkin 2011), and is becoming even more central as we move into the Fourth Industrial Revolution (4IR) (Schwab 2017). The 4IR is characterized by a fusion of technologies, which is blurring the lines between the physical, digital, and biological spheres.

Increasingly, systems and applications operate in a context in which the flow of information has direct bearing on daily lives of billions of people, where two fundamental characteristics of the use of such information emerge – Transparency and Privacy. Transparency is the quality that allows participants of a community to know what the particular processes and agents are being used in its functioning. It is generally regarded as a means to enable checks and balances within this community, ultimately providing a basis for trust among participants of that community. When the community is regarded as being the entire society, these checks and balances are reflected in its political system to prevent misuse by any of the parties involved.

One of the mechanisms created to increase transparency in political systems is the enactment of regulations ensuring the right of its members to access to information in a variety of contexts, ranging from government-produced

[1] *https://googleblog.blogspot.com/2012/05/introducing-knowledge graph-things-not.html*

[2] For instance, Wordnet (Miller 1995), DBPedia (Lehmann et al. 2013), Yago (Suchanek et al. 2007), CYC (Lenat 1995), NELL (Carlson et al 2007), ConceptNet (Speer and Havasi 2010)

ISSN: 1839-4183

information and data to consumer-related information regarding goods and products, as well as the right of individuals to freely create, publish and access information. As we will explain later, regulations do not comprehend only legislation, i.e. enacted laws and statutes, but all kinds of legal instruments of governance—hard law, soft law, policies and ethics.

The free flow of information, on the other hand, may conflict with another basic human right, that of Privacy (Universal Declaration of Human Rights 1948, art. 12). There are many definitions for Privacy (Paci, Squicciarini, and Zannone 2018), but in essence they all refer to the right of an individual to control how information about her/him is used by others. Data Protection and Privacy have had, and still have, different regulatory regimes that are applied in a variety of jurisdictions in both Civil and Common Law legal systems. For instance, Privacy is considered a fundamental (human) right by the recent General Data Protection Regulation (GDPR) enacted in Europe in May 2018. It is also assumed as a constitutional right in many EU countries. But this is not so in the USA, in which privacy and data protection do not qualify as specified fundamental rights under the Constitution.

The situation in Common Law countries is nuanced, varying with some subtlety between different national jurisdictions. For instance, there is no general law right to privacy in Australia. Although Australia is a signatory to the International Covenant on Civil and Political Rights (adopted by UN General Assembly on 16 December 1966, and in force from 23 March 1976), the international law right to privacy conferred under Article 17 of the ICCPR has not been enacted into Australia's domestic law (Watts and Casanovas 2018). The Privacy Act 1988 still regulates information privacy in the Commonwealth public sector and the national private sector. It covers personal information and sensitive information (such as health information, ethnicity, sexual preference, trade union membership). This situation is also evolving, partially fuelled by the palliative reaction against Covid-19. From 2019 the Federal Government has been engaged in several trends to update the definition of personal information.[3] The recent Issues Paper (Australian Government 2020) on the reform displays a wide set of issues, including whether a statutory tort for serious invasions of privacy should be introduced, and whether the Privacy Act should include a 'right to erasure'.

It is our contention that these and similar needed reforms around the world should consider the technical developments occurring in computer science and semantic web studies. In order to deal with the myriad of often conflicting cross-cutting concerns, Internet applications and systems should incorporate adequate mechanisms to ensure compliance of both ethical and legal principles. In order to be effective, we claim that the use of Knowledge Graphs ought to provide support for these concerns—trust, privacy and transparency. In this paper we propose a framework that enables this support. In Section 2 the regulatory framework approach and some concepts stemming from the work already done in the field will be defined. It will also describe the preliminary legal background. Section 3 backgrounds three main general concepts—trust, privacy, and transparency—that will be taken into account in our modelling. Section 4 presents an illustration by way of example, and Section 5 discusses the proposed KG representation framework, showing its application in the illustration. Section 6 reflects on the previous example, and Section 7 draws some conclusions and points to future work.

2. A LEGAL GOVERNANCE APPROACH

In this section, we argue that the privacy legal regulatory framework does not solely lean on national legislation but encompasses other legal instruments of transnational nature (such as protocols, standards, best practices and ethical principles and values). The field of privacy is specially suitable for a broader legal governance, in which hetero-regulatory, co-regulatory and self-regulatory instruments tend to coexist in different ways according to the specific contexts created by the normative systems at stake in regional, national, international and transnational economic, social and political spaces (Pagallo et al. 2019; Pagallo, Casanovas and Madelin 2019).

Legal governance can be defined as the processes and practices of implementing the set of normative systems put in place in specific contexts for a variety of scenarios; i.e. the process of creating *sustainable legal ecosystems*. From this standpoint, legality is the result of the coordination of different types of agency (artificial and/or human,

[3] *I.e. issuing several Recommendations for reform (Australian Government 2019). Among others (i) updating the definition of 'personal information' to capture technical data and other online identifiers (Recommendation 16(a)); (ii) strengthening existing notification requirements (Recommendation 16(b)); (iii) strengthening consent requirements and pro-consumer defaults (Recommendation 16(c)); (iv) and introducing a direct right of action to enforce privacy obligations under the Privacy Act (Recommendation 16(e)).*

using socio-technical, cognitive-socio-technical systems or normative Multi-agent Systems). Some legal instruments—such as rules extracted from legal norms—can be automated. Others, for instance, ethical values and principles or best practices regarding the monitoring of regulatory systems, cannot be fully hardcoded, as they require human intervention and decision-making (Koops and Leenes, 2014) This is the case with trust, transparency, privacy and data protection. Their implementation requires building institutions to hold complex models of legal governance. But this is not saying that they cannot be semi-automated. On the contrary, from this perspective, appropriate automation can facilitate human control and monitoring.

Accordingly, privacy and data protection can be considered from the legal governance approach. This means that they can be implemented through the construction, development and implementation of a technological toolkit, comprising data mining, data analytics and the linked open data tools of the later developments of the Semantic Web—i.e. the Web of Data. The Knowledge Graphs approach that we embrace in this article is related to this dimension.

2.1 COMPONENTS OF LEGAL GOVERNANCE

Since 2010, many large-scale RDF datasets have been created. For instance, in 2017, Freebase 1 had 2.5 billion triples;[4] DBpedia2 had more than 170 million triples. LOD (the Linked Open Data cloud) connects more than 3000 datasets, with more than 84 billion triples. The number of data sources doubles every three years (Zou and Özsu 2017). This has created a data space, strongly interconnected, in which law and government can have a leading role, as legal documents and, mostly, legal content and knowledge, are increasingly offered for public consumption (Casanovas et al. 2016). Linguistic resources for the legal domain are increasingly being identified, classified and annotated (Martín-Chozas et al. 2019, Rehm et al. 2020). Ontolex-lemon, the vocabulary for lexical resources in the Web of Data, have been extended to create databanks of legal terminologies that can be used automatically by dictionaries (Rodríguez-Doncel et al. 2015).

However, new legal issues arise, such the use by LOD of crowdsourced vocabularies, where there is no authority imposing one interpretation over another. There is no evidence so far of case-law nor out-of-court disputes regarding linked data resources. On the other hand, as is well known, from 2000 onwards many data breaches have been reported—AOL, NETFLIX, Equifax, Cambridge Analytica etc.[5] — and the lawsuits that followed indicate a certain alarm about a general state of uncontrolled surveillance (Norris et al. 2017, Zuboff 2019). Thus, the legal dimensions of security, intellectual property, patents, licenses and, especially privacy and data protection should be taken into account and applied to the use of legal resources and the building of new tools for the web of data (Rodríguez-Doncel et al. 2016).

Legal resources should be differentiated from legal sources. The former refers to the large number of existing legal vocabularies and documents on the Web of Data. The latter refers to the specific content that 'counts as legal' at regional, national, international and transnational levels to be effectively implemented or enforced. Determining what is 'valid' law, what *counts as legal*, is in itself a non-trivial theoretical operation that is usually performed through the concepts of doctrine, legal theory, and checks-and-balances (Sartor 2005, Peczenik 2011). For our purposes, it is worth noting that a taxonomy of a legal quadrant, or *legal compass* (Fig. 3) would suffice (i) to classify and annotate a variety of sources (ii) that are deemed necessary to produce the 'ecological validity' of a regulatory system—i.e. the condition of sustainability of legal ecosystems for a cluster of stakeholders (Poblet, Casanovas and Rodríguez-Doncel 2019). This legal compass (i) reflects and endorses the two sides of the rule of law (binding power and social dialogue), (ii) and assumes as

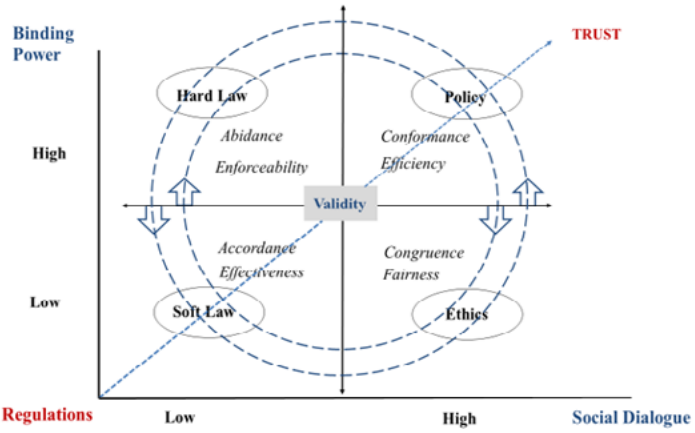

FIGURE 1. Legal compass for the rule of law. Source: Poblet, Casanovas and Rodríguez-Doncel (2019)

[4] A triple is a set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions.

[5] See a short list of the most notorious cases at https://en.wikipedia.org/wiki/Data_breach

ISSN: 1839-4183

a pre-requisite that they can be partially coded through semantic representations.

In the scheme of Fig. 1, institutional strengthening and trust are intended and eventually produced through a variety of sources that must be ordered beyond a determined threshold to build valid regulatory systems. Validity, and specifically legal ecological validity, emerges from the degree of compliance with several requirements—e.g. enforceability, efficiency, effectiveness, fairness (justice)—that make the system acceptable and sustainable. Thus, validity qualifies as a second order property that encompasses the whole regulatory system (not only a specific rule or norm). This approach is also compatible with recent surveys on business languages and compliance (Hashmi 2018) but shows that legal compliance requirements can be more complex than those set by regulatory languages.

2.2 A PRELIMINARY FRAMEWORK

As already stated, there are many differences between the USA and EU approaches for regulating privacy. Whereas EU laws consider 'privacy' as a human right, a constitutional fundamental right, and a fundamental EU right under the Article 7 and 8 of the EU Charter of Fundamental Rights (2000 and 2012) (González-Fuster 2014, Blasi 2016),[6] US "values it as a liberty over and against the state" (Blasi 2014). Thus, data protection of personal data should be differentiated from privacy *tout court.* In contrast, as noticed by many scholars, the United States does not provide for an overall legal expectation of privacy. The collection and processing of personal data is regulated based on the type of data at stake. Thus, data related to healthcare is subject to the Health Insurance Portability and Accountability Act (Kennedy–Kassebaum Act, 1996) commonly known as HIPAA, and financial data is governed by the Financial Services Modernization Act (Gramm-Leach-Bliley Act, 1999), known as GBLA.[7]

Legal requirements might be quite detailed, and different according to national and jurisdictional frameworks. Related to privacy, to link the knowledge graph approach to legal governance, we will endorse a broader conceptual stance, linking the main concepts to be modelled (i) to the middle-out approach defined by Pagallo et al. (2019a, 2019b) and (ii) to the existing data protection engineering approaches.

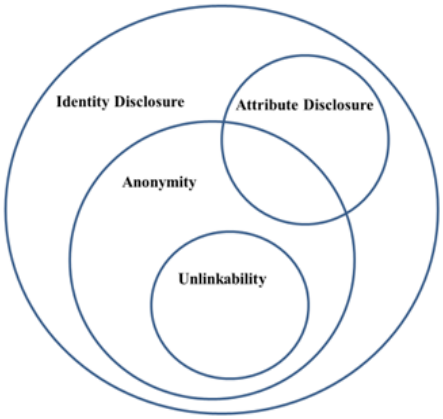

FIGURE 2. Relationship between Anonymity, Unlinkability and Disclosure. Source: Torra (2017, p. 9). Quoted with permission

Following Pfitzmann and Hansen (2010), Torra (2017) organized their terminology for data minimization into four interrelated categories, depicted in Fig. 3 as Venn diagrams. Henceforth, Torra extended and refined their original data minimization strategy.

Pfitzmann and Hansen (2010) originally provided the following three definitions: (i) Anonymity of a subject means that "the subject is not identifiable within a set of subjects, called 'anonymity set'"; (ii) From an adversary (intruder, attacker…) perspective, anonymity of a subject means that "the adversary cannot achieve a certain level of identification for the subject s within the anonymity set"; (iii) Unlikability of two or more items of interest (IoI) from an attacker's perspective means that "within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether those IoIs are related or not.[8] Torra (ibid. 10) provides three additional ones: (iv) Disclosure, that "takes place when attackers take advantage of the observation of available data to improve their knowledge on some confidential information about an IoI"; (v) Identity Disclosure, "when the

[6] *The Charter of Fundamental Rights of the EU Union (2000-2012) states in Art. 7: "Everyone has the right to respect for his or her private and family life, home and communications". Art. 8.1 reads: "Everyone has the right to the protection of personal data concerning him or her", and Art. 8.2: "Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified."*

[7] *https://www.hipaajournal.com/comparison-of-european-and-american-privacy-law/*

8 It is worth noting that unlinkability is deemed a sufficient non-necessary condition, as it implies anonymity. However, Torra (2009, 9) points out that there might be a case in which linkability is possible, but anonymity is not.

TABLE 1: Fair Information Principles Practices. Source: Langheinrich (2001).

1. *Openness and transparency*	There should be no secret record keeping. This includes both the publication of the existence of such collections, as well as their contents.
2. *Individual participation*	The subject of a record should be able to see and correct the record.
3. *Collection limitation*	Data collection should be proportional and not excessive compared to the purpose of the collection.
4. *Data quality*	Data should be relevant to the purposes for which they are collected and should be kept up to date.
5. *Use limitation*	Data should only be used for their specific purpose by authorized personnel.
6. *Reasonable security*	Adequate security safeguards should be put in place, according to the sensitivity of the data collected.
7. *Accountability*	Record keepers must be accountable for compliance with the other principles.

adversary can correctly link a respondent to a particular record in the protected data set"; (vi) Attribute Disclosure "when the adversary can learn something new about an attribute of a respondent, even when no relationship can be established between the individual and the data".

These concepts stand on the shoulders of the *Fair Information Practice Principles* (FIPPs) proposed by US Secretary's Advisory Committee on Automated Personal Data Systems in a 1973 Report, *Records, Computers and the Rights of Citizens.* This Report was followed by the US Privacy Protection Study Commission Report on *Personal Privacy in an Information Society* (1977). Legal experts—Alan Westin (1967)—, AI and law experts—such as Layman E. Allen—, and computer scientists—Willis H. Ware (RAND Corporation)—were involved in their development. Table 1 summarises the so-called FIPPs.

In 2004, Kim Cameron, Chief Identity Architect of Microsoft, wrote and blogged what he would call the "7 Laws of the Internet". His approach was to set a *metasystem identity layer,* i.e. "to develop a formal understanding of the dynamics causing digital identity systems to succeed or fail in various contexts, expressed as the Laws of Identity. Taken together, these laws define a unifying identity metasystem that can offer the Internet the identity layer it so obviously requires.[9]

In the following years, Ann Cavoukian (2006, 2007) aligned this Internet Identity Metasystem Layer with FIPPs. The result was the proposal of Privacy by Design

TABLE 2. Privacy Design Strategies. Source: from Hoepman (2014) and ENISA (2015)

Privacy Design Strategies [J.H.Hoepman, 2014]			
Data-Oriented Strategies	**Minimize**: The amount of personal data that is processed should be restricted to the minimal amount possible.	Select before you collect. Anonymization. Use of pseudonyms	**Design Patterns**
	Hide: Any personal data, and their interrelationships, should be hidden from plain view.	Encryption, Mix networks, Attribute base credentials, Anonymization. Use of pseudonyms	
	Separate: Personal data should be processed in a distributed fashion, in separate compartments whenever possible.	No specific design patterns known	
	Aggregate: Personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful.	Aggregation over time (used in smart metering), Dynamic location granularity (used in location based services), and k-anonymity	
Process-Oriented Strategies	**Inform**: Data subjects should be adequately informed whenever personal data is processed.	Platform for Privacy Preferences (P3P), Data breach notifications, Patterns from the Human Computer interfacing perspective	
	Control: Data subjects should be provided agency over the processing of their personal data.	No specific design patterns known	
	Enforce: A privacy policy compatible with legal requirements should be in place and should be enforced.	Access control, Sticky policies, Privacy rights management (a form of digital rights management involving licenses to personal data)	
	Demonstrate: Be able to demonstrate compliance with the privacy policy and any applicable legal requirements.	Privacy management systems , the use of logging and auditing	

(PbD), the process for embedding privacy principles into design specifications architectures. However, computational modelling, the specific engineering paths to make it happen was—and still is—a model to be assembled or to be built. Hoepman (2018) puts it in the following way:

"Privacy by design is a system development philosophy that says that privacy should be taken into account throughout the full system development lifecycle, from its inception, through implementation and deployment,

[9] *"We need a unifying identity metasystem that can protect applications from the internal complexities of specific implementations and allow digital identity to become loosely coupled. This metasystem is in effect a system of systems that exposes a unified interface much like a device driver or network socket does. That allows one-offs to evolve towards standardized technologies that work within a metasystem framework without requiring the whole world to agree a priori." Cameron (2004).*

ISSN: 1839-4183

all the way until the system is decommissioned and no longer used. In software engineering terms this makes privacy, like security or performance, a software quality attribute or non-functional requirement."

Hoepman (2014) represented the existing engineering strategies (see also ENISA 2015) as design strategies that describe a fundamental approach to achieve a certain design goal. Thus, a *privacy design strategy* could be understood as a "design strategy that achieves (some level of) privacy protection as its goal". The PbD analytical framework can be mainly set through eight strategies: minimize, hide, separate, aggregate, inform, control, enforce and demonstrate (Table 2). The strategies are assembled into two classes: (i) the first four are data-oriented strategies and (ii) the latter four are process-oriented strategies. The European Union Agency for Network and Information Security (ENISA) embraced this approach in 2015, as there was an increasing need to create a common framework for policy makers, legislators, service providers, data protection authorities and standardization bodies. It is similar to the framework eventually assumed in GDPR provisions. Colesky, Hoepman and Hillen (2016) correlated and map the strategies against privacy patterns, adding the notion of 'tactics' as an additional level of abstraction between them.

Nevertheless, this remains an open field. We have already quoted the extension of the Pfitzmann and Hansen (2010) minimization strategy by Torra (2017). There are many other possibilities. For instance, Rastogi, Gloria and Hendler (2017) have developed a general method to enhance privacy in the cloud, deploying mobile applications dynamically on a scalable on-demand hardware and software platform. From a legal governance perspective, Casanovas et al. (2014) have proposed an indirect strategy for security platforms in which coding should be combined with the construction of ad hoc 'anchoring' institutions to monitor and control the outcomes of the information processing flow.

In the following sections, we illustrate how these considerations can be supported using Semantic Web formalisms to represent the relevant dimensions. Rather than a final solution, we show the essential aspects that need to be addressed in any representation that is incorporated in Knowledge Graphs to support the regulatory framework around Privacy and Transparency.

3. BACKGROUND CONCEPTS

Before detailing our framework, we briefly present our definition for each of the basic concerns. We have detailed each of them in other publications, as referenced in each sub-section.

3.1 TRUST

The issue of trust has been prevalent in the Internet since its popularization in the early 90s—see the Grandison and Sloman (2000) survey. Attention has focussed on the lower layers of the Internet Architecture, emphasizing authentication, which deals primarily with verification of identity. More recently, with the advent of the Web and social networks, the cybersphere and society generally, have become heavily influenced by information (and misinformation) that flows in news sites and social networks in the Internet. There are many studies carried out in several disciplines attempting to characterize and understand the spread of information in the cybersphere, as well as how this affects society—see Marwick and Lewis (2017) for an overview.

The original vision for the Semantic Web included a "Trust" layer, although its emphasis was more on authentication and validation with static trust measures for data. There have been many efforts in representing trust, including computational models.[10] As proposed initially in Almendra and Schwabe (2006), and later in Laufer and Schwabe (2017) and Schwabe, Laufer and Busson (2019), the approach used here is based on the work of Gerck (1997) and Castelfranchi and Falcone (2001), taking the view that trust is "knowledge-based reliance on received information", that is, an agent (i.e., a person or a software program) decides to trust (or not) based solely on her/his knowledge, and the decision to trust implies the decision to rely on the truth of received or on already known information to perform some action.

[10] A general survey can be found in Pinyol and Sabater-Mir (2013); Artz and Gil (2017) presents an excellent earlier survey for the Semantic Web; and Sherchan, Nepal, and Paris (2013) surveys trust in social networks. In the Linked Data world, it is clear that facts in Semantic Web should be regarded as claims rather than hard facts (Bizer and Cyganiak 2009), which naturally raises the issue of trust on those claims.

In terms of a Knowledge Graph, an agent wishing to perform an action must first filter those information items it deems "trusted", i.e., it will use them to perform the intended action. Since it is not possible to "half-act", in this sense trust is binary—either the agent trusts the information, or it does not. A more extensive discussion can be found in Schwabe, Laufer and Busson (2019). Trusted information as the basis for supporting privacy and transparency is discussed next.

3.2 PRIVACY

As discussed in section 2.2, there are many definitions for privacy.[11] Tavani (2007) classified privacy theories into four broad types: the nonintrusion, seclusion, limitation, and control theories. For our research, we adopt the Restricted Access/Limited Control (RALC) Theory proposed by Moor and Tavani (2001). RALC presupposes that an adequate theory of privacy needs to differentiate the concept of privacy from both the justification and the management of privacy. Accordingly, the RALC framework has three components: an account of the concept of privacy, an account of the justification of privacy, and an account of the management of privacy. "RALC requires that one must have protection from intrusion and interference and information access, it addresses concerns not only about protecting informational privacy (as described in the control and the limitation theories) but also about protection against the kinds of threats described in the nonintrusion and the seclusion theories as well." (Tavani 2007, 10).

Based on this theory, we define Privacy as "controlled access to information related to an agent". In order to ensure privacy, it is necessary to answer three questions:

- Q1: What types of Actions are allowed (and controlled) over Knowledge Items (KIs)?
- Q2: What are the relation types between some Agent and a KI that entitle this Agent to establish a Privacy Rule governing Actions over that KI?
- Q3: How to resolve conflicts between applicable rules?

3.3 TRANSPARENCY

Generally speaking, according to Meijer (2013, 430) and others (Schudson 2015), transparency can be defined as "the availability of information about an actor that allows other actors to monitor the workings or performance of the first actor." It contemplates the "capacity of outsiders to obtain valid and timely information about the activities of government or private organizations".[12] Transparency presupposes the involvement of an observed and an observer (Berstein 2017). In contrast to privacy, that is concerned with information about individuals, transparency concerns any type of information, although it may make a difference if the producer (author) is an organization or an individual (Heimstäd and Dobusch 2018).

Transparency is also related to privacy. In the so-called "Attention Economy" (GDPR, CE 2016), for example, the information about users and consumers are a primary source of value, and companies actively seek to obtain as much information about users as possible. This can be in direct conflict with privacy rights of users, who have the right to control how the information about them is used by others. In this state of affairs, transparency can support the resolution of potential conflict with privacy policies: the disclosure of information about the company's processes and procedures associated to the individual can contribute to foster trust, showing that it is compliant with these regulations and thus to stimulate authorization on the part of the users. This has been subject to recent regulations such as the General Data Protection Regulation enacted by the European Union.

From a more abstract approach, both Privacy and Transparency relate to controlling actions over information, and who can define such controls. As such, they can be regarded as two points in the same control dimension. Privacy tends to limit or restrict actions over information items, whereas Transparency tends to allow (in some cases, mandate) actions over them, which explains the natural tension that exists between the two.

[11] *See for example Smith, Dinev and Xu (2011). Paci, F., Squicciarini, A., Zannone (2018) and Such and Criado (2018) present surveys on multi-party privacy.*

[12] *https://www.britannica.com/topic/transparency-government#accordion-article-history*

ISSN: 1839-4183

4. AN ILLUSTRATION - DISASTER RELIEF DONATION

The goal of this example is to illustrate the complex and interdependent nature of trust, privacy and transparency; it will later be used to show the expressive power of the framework and how it incorporates the various concerns at play.

Consider a scenario in which a disaster has occurred, such as the fire that burned the roof of the Notre Dame Cathedral, or the bush fires in Australia. In response to several demands, Ed wants to donate some money to help with the disaster relief actions. Ed has received several donation requests from different organizations and needs to choose one of them to make the donation. However, given his past experience, Ed wants to make sure these are legimate request as opposed to frequently occurring scams. In other words, he wants to make sure that the donated money will actually be used for the relief actions, rather than being misused, e.g., funding the organization's basic infrastructure, or employed in another action, or even pocketed by unscrupulous officers of the receiving organization.

Ed formulates a rule saying that he only trusts organizations that openly publish who are their financial officers, and their financial records. Financial records must be validated by accredited audit organizations. Furthermore, because of personal reasons, Ed does not want to contribute to an organization in which George is an officer. This rule can be regarded as an application-related rule, akin to what is referred to as a "business rule" in traditional software development.

Let us assume that there is a law that stipulates that not-for-profit organizations must publicly identify their officers and that George is an officer of *ReliefOrg*, an NGO dedicated to raising funds for and helping disaster relief efforts. George, being a very reserved person, has a privacy rule that stipulates that his association with any organization, including *ReliefOrg*, should not be made public.

Ed receives a request for donation from *ReliefOrg* and needs to decide whether he should donate or not. The first step Ed follows is to verify if the officers of *ReliefOrg* are published in the KG.[13] Here we can see a potential conflict between transparency and privacy rules. George's relation with *ReliefOrg* should be accessible in the KG, according to not-for-profit legislation. On the other hand, George's privacy rules would prevent this access. Since transparency legislation in this case has higher precedence than personal rules, this association can then be used by Ed's rules, and therefore *ReliefOrg* would not be accepted as a recipient of donations by Ed.

Consider now a slightly different scenario where George is not an officer of *ReliefOrg*, and Ed wants to check the financial integrity of *ReliefOrg*. He retrieves the financial report for *ReliefOrg* from the KG, but wants to make sure it has been audited, so he looks for a certification of the financial report. He finds out it has been audited by *AuditInc*, which is unknown to him, so he needs to verify that it has an accreditation certificate from a public authority. If such a certificate is available, Ed analyses the financial report finding nothing wrong in principle.

However, tipped by a friend, Ed learns that George may in fact be one of the owners of *AuditInc*. He then checks *AuditInc* to see if its owners are listed, and whether George is one of them. Given George's privacy rule plus the fact that *AuditInc* is not a not-for-profit organization, his privacy rule would prevent access to the owner relationship, so Ed would not know that George is one of the owners, thus deciding to contribute. Note that in this scenario, George's privacy rule would apply not only to *AuditInc*'s information, but also to information furnished by others, for example, a photo in a social network where George appears in the annual Christmas party of *AuditInc* with a caption mentioning his role as one of the partners, or perhaps with a badge on his neck identifying him as such.

We next describe our proposed KG Usage framework and subsequently analyse this example showing how it can be represented by it.

5. A SUMMARY OF THE KG USAGE FRAMEWORK.

Figure 3 shows a diagram of the use of information within a KG. "Using a KG" is represented as a Request made by some Agent for an Action over a Knowledge Item (KI). We assume the existence of an underlying RDF graph, which would be equivalent to the "traditional" definition of KGs.

[13] See for instance https://permid.org or https://opencorporates.com as examples of KGs with this type of information.

An RDF graph is simply a collection of statements, each in the form of a triple <subject, property, object> such as <Australia, capital-city, Canberra>. An RDF named graph is one for which there is an identifier to refer to it.[14] The actual KG is formed by defining several named graphs over this underlying RDF graph as a way to structure the RDF triples into Knowledge Items, similarly to Items, Statements and Qualifiers in Wikidata (Erxleben et al. 2014). Thus, a Knowledge Graph represents a collection of interlinked descriptions of KIs – real-world objects, events, situations or abstract concepts – where:

- Descriptions have a formal structure that allows both people and computers to process them in an efficient and unambiguous manner;
- Entity descriptions refer to one another, forming a network, where each entity represents part of the description of the entities related to it.[15]

We propose to represent the KG as a collection of Knowledge Items (KIs), each of which as a nanopublication[16] (Groth, Gibson and Velterop 2010). A nanopublication "offers a supplementary form of publishing alongside traditional narrative publications", consisting of three parts representable by RDF graphs: "(i) an assertion (a small, unambiguous unit of information), (ii) the provenance of that assertion (who made that assertion, where, when, etc.), (iii) the provenance of the nanopublication itself (who formed or extracted the assertion, when, and by what method)." (Golden and Shaw 2016)

5.1 KNOWLEDGE ITEM REPRESENTATION

A KI, as all nanopublications, comprises an assertion graph, a provenance graph and a publication info graph.

The assertion graph of a KI contains a set of assertions about its content. The assertions in this graph are a subset of the assertions in the underlying RDF graph. As an example, if a KI refers to an event, the assertion graph would contain statements about the participants and their roles, location information, date information, depictions (photos, videos, …) and so on.

The provenance graph of the nanopublication will contain provenance information about the assertions in the assertion graph (e.g.; what image or natural language processing software was used, recorded location info, whether the assertions were inferred using some inference engine, etc.). The provenance graph can be used to represent, to the desired level of detail, the supporting information for the assertions. For example, if an automated face recognition algorithm was used, the provenance information represented in the provenance graph of the nanopublication may inform which algorithm was used, which parameters were used in this particular case, and a confidence factor of the algorithm about the correctness of the extraction. Another use of provenance can be seen in the case of a statement stating that, for example, <Barack Obama> placeOfBirth <Hawaii>. The provenance information may include documentation to support its truthfulness, such as a reference to a birth certificate that states that indeed the place of birth of Barack Obama is Hawaii.

The publication info graph will contain metadata about

```
Define EvalRequest(Agent, Action, KI),
    1. RS <- UsageRuleSet(KI).
    2. Let Aut <- EvalUsageRuleSet(RS, Agent, Action, KI)
    3. If Aut = "allowed", Let TS <- TrustRuleSet(Agent); Let TGA <- Eval-
       TrustRuleSet (TS, Agent, Action); Execute (Action, KI, TGA).
Define EvalUsageRuleSet (RS, Agent, Action, KI}
    4. Let RS <- Sort-by-Precedence(RS, decreasing).
    5. Let A <- DefaultAuthorization.
    6. For each R in RS,
        a) Let AR <- EvalRule(R, Agent, Action, KI);
        b) If AR = "Allowed" or AR = "Denied",
            return AR.
    7. Return A.
```

FIGURE 4. Request Evaluation Algorithm

the creation of the KI itself (as opposed to the information contained in its assertions sub-graph), such as its author, creation date, etc…

[14] https://www.w3.org/TR/rdf11-concepts/#section-rdf-graph

[15] https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/

[16] http://nanopub.org . "A scholar can promote small pieces of information within her work using the practice of nanopublication. Nanopublications include useful and usable representations of the provenance of structured assertions. These representations of provenance are useful because they allow consumers of the published data to make connections to other sources of information about the context of the production of that data." (Golden and Shaw 2016)

ISSN: 1839-4183

5.2 CONTROLLING USAGE

In our framework, Privacy and Transparency refer to Request for actions over some information (in a KI), for which an Authorization must be granted, according to the Rules set forth by the relevant stakeholders. Stakeholders include persons "related" to the KI, as well as institutional agents such as "the State" (whose rules are stated as laws). Rules may be based (make use of) on any information available in the KG. We refer to both Privacy and Transparency rules collectively as Usage Rules.

Figure 4 introduces the basic algorithm to evaluate a Request. The first step in granting an Authorization for the action is to determine which KIs in the KG are trusted by, the Agent requesting the Action. This is realized by a Trust Engine that collects applicable rules and includes them in a Trusted sub-graph of the KG. We refer to the set of trusted KIs by Agent α as the Trusted KG(a).

Once the Trusted KG has been determined, a set of usage rules are evaluated, which entails determining applicable rules (UsageRuleSet), evaluating them (EvalUsageRuleSet) and resolving conflicts if they arise. In order to evaluate a usage rule, it is also necessary to evaluate the trust rules of the author of that usage rule, which needs to be based on trusted KIs. Thus, a set of Trusted KG (ai) is defined, one for each Author ai of an applicable usage rule. The result of this process defines the final Authorization for the Action in the Request that was made. If the Authorization is "Allowed", the Action is carried out using the TrustedKG of the Agent that requested the Action.

Stated in words, the algorithm stipulates that, to evaluate a request for an Action by an Agent over a Knowledge Item (KI), the following steps should be performed (line numbers correspond to those in Figure 4).

1. Collect the applicable rules referring to the KI and store them in variable RS
2. Apply the collected rules in RS for the requested Action by Agent on the KI.
3. If the Action was allowed, then collect the trust rules for Agent and store them in variable TS, then use these collected rules to find out the set of trusted information items by Agent in the KG (stored in TGA), and perform the requested Action on the KI, using the trusted information in TGA

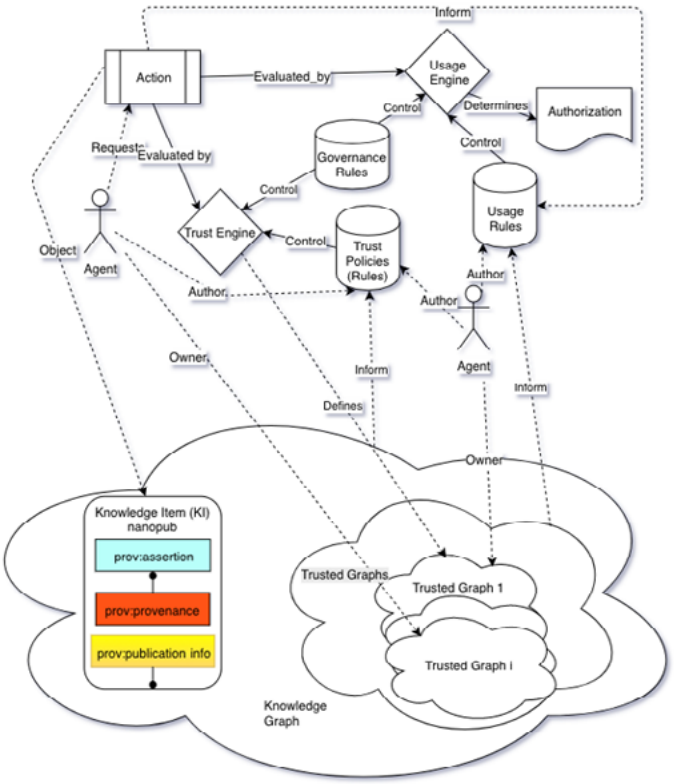

FIGURE 5. Usage process for information in a KG

In order to apply a set of usage rules RS for a request for an Action on a KI by an Agent (as specified in step 2. above), one must perform the following steps:

4. Sort (reorder) the rules in RS in decreasing order of precedence (i.e., the rule with highest precedence comes first)
5. Assume there is a default authorization – i.e., if nothing is said, an Action is allowed (or denied) by default.
6. Consider each rule R in the set RS in turn. For each one,
   - Determine the authorization value by applying this rule R for the Action on the KI requested by Agent.
   - If the authorization value can be determined, then return it as the result of the algorithm. It is possible that the relevant rules are unable to determine an authorization value, e.g., because of insufficient information in the KG.
7. If no value could be determined, return the default authorization as the result of the algorithm.

To determine the final Authorization value, a conflict resolution strategy must be employed, which is in turn subject to Governance Rules. This algorithm abstracts the essential decisions that must be made, to wit:

1. Who can formulate a rule for a given KI? – in line 1;
2. How are conflicts between rules resolved? – in line 5-8
3. What are the allowed actions over KIs? – in line 4.

We detail possible answers to these questions in the following sub-sections.

5.3 RULES

The UsageRuleSet(KI) function call determines what are the applicable rules given a KI. For trust rules, it is determined by the author of the request. For privacy rules this corresponds to answering the question "who has the right to define a Privacy Rule that controls actions over this KI?" The definition of Privacy itself indicates that it must be any agent that is somehow related to the information contained in the KI. Any useful instantiation of the framework must spell out what are the accepted relation types, which can include:

- An identification property for any agent that is included in the KI – for example, some person appearing in a posted photo or video;
- Any relation denoting referral to a person included in the KI – for example, some person cited in a post;
- Any creation or authorship relation;
- Any agent related to the creation of the KI – for example the author of a video posted by someone else;
- Any agent who has legal jurisdiction over an agent identified or mentioned in the KI;
- The Agent representing the legal system(s) that has(have) jurisdiction over the KG, over the Agent or over the Action request.

The presence of such relations can directly occur in the KG (i.e., as a typed edge), or as a composition of valid relations. Furthermore, the rules may (or may not) allow the use of inferred relations in the KG.

Rules are of the form antecedent => consequent, both of which are sets of statements (Almendra and Schwabe 2016). The antecedent of privacy rules may refer to any statements in the KG, including

- Any statements in the subgraphs in the KI's nanopublication (assertions, provenance and publication info);
- The identity of the agent requesting permission;
- The type of Action;
- Information in the KG serving as contextual information, such as
  - Date/time of the request;
  - Current location of agents involved;

Seen as nanopublications, the actual specification of the rule is given in its assertion graph, using a notation such as N3Logic (Berners-Lee et al. 2006) or SWRL.[17] Governance Rules are Rules that include other Rules (in either of antecedent and consequent), so in this sense they are meta-rules – i.e. rules about rules.

5.4 CONFLICT RESOLUTION

The Sort-by-Precedence (RS, decreasing) function call sorts the enabled rules in descending precedence order, typically combining several kinds of information to establish order relations among rules.

Some possible complementary order relations that can be employed are:

- Hierarchical relations between users – rules defined by a higher-ranked user take precedence over rules defined by lower-ranked ones. For example, rules established by laws (authored by state agents) take precedence over rules stated by individuals (common citizens);
- Hierarchical relations between relation types. Rules defined by users related to the KI through a higher-ranked relation type take precedence over rules defined by users related via a lower-ranked relation type. For example, one may state that relation type "identifies" takes precedence over type "mentions". Thus, in a video where a person A is identified, and person B is mentioned in a conversation, person A's rules would take precedence over person B's rules.

[17] *https://www.w3.org/Submission/SWRL/*

ISSN: 1839-4183

Since hierarchies are partial orders (which means that there may be items without an ordering between them), they may not define precedence completely, so further conflict resolution strategies are still needed. Such and Criado (2018) identified six categories of strategies that can be employed. Most require user involvement at runtime, but the aggregation-based class of strategies can be easily incorporated into an algorithm. Strategies in this class define an aggregation function such as consensus, majority, minimum fixed number of votes, permit-overrides, deny-overrides, etc, and replace the set of conflicting rule results by a single aggregated result.

An alternative to the aggregation approach is to decompose the KI into finer-grained elements so that each one is subject to only a single rule. This makes sense when the Action to be performed can be stated as the composition of the same operation performed on each element independently. For the elements where the Action is denied, the resulting composite object is altered. For example, in a group photo showing several people, some may allow publication other may deny publication. In this case, it would be possible, to blurr out the face of the persons who have denied publication.

6. EXAMPLE SCENARIO REVISITED

In this section we re-examine the illustrative example scenario in light of the KG usage framework, showing how the most important aspects are represented. The first rule captures Ed's requirement about Non-Profits: he only trusts organizations that openly publish who are their financial officers, and their financial records. Financial records must be validated by accredited audit organizations.

The trust rule below captures this, expressed using N3Logic with extensions. We state under KG some statements we assume to be present in the KG:

KG

NonProfit subClassOf org:Organization. AuditCo subClassOf org:Organization.

CertificationAgency subClassOf org:Organization. Donate subClassOf Action.

<*ReliefOrg*> type NonProfit. <*AuditInc*> type <CertificationAgency>.

<*ReliefOrg*> officer <George>. <*AuditInc*> officer <George>.

RuleEd1

{?O type org:Organization; hasOfficial: ?Ofc; hasFinancialRecord ?FR. ?Ofc type foaf:Person.

?FR assertions ?FRa. ?FRa log:semantics ?FRaS; ?FRAS log:includes

{ ?FR auditedBy ?Aud. ?Aud type AuditCo}.

?FR provenance ?PFr. ?PFr log:semantics ?PFrS. ?PFrS log:includes {prov:hasPrimarySouce ?DOCS}. ?DOCS assertions ?DOCSa. ?DOCSa log:semantics ?DOCSaS.

?DOCSaS log:includes {AuditCo certifiedBy ?CA. }. <TrustedGraphEd> author <Ed>

log:semantics ?TGEd. ?TGEd log:includes {?CA type CertificationAgency}}

=> {<TrustedGraphEd> :add {?O type:NonProfit}. }

Stated in plain language, the KG contains statements affirming that NonProfit, AuditCo and CertificationAgency are a kind of Organization, and Donate is an Action. It also affirms that *ReliefOrg* is a NonProfit organization; *AuditInc* is a Certification Agency and that George is an officer of both *ReliefOrg* and *AuditInc*.

RuleEd1 reflects Ed's trust criteria – Ed's Trusted Graph will include a NonProfit organization only if its financial records include a list of its officers, and the financial record itself is audited by a certified Certification Agency.

This is a trust rule since it refers to KIs which are needed as input for Ed to allow taking the "Donate" action.

Ed's second rule is in fact a business rule – he does not want to contribute to an organization in which George is an officer. While this could be embedded in application code using the KG, we express it here as a rule as well, to facilitate the discussion.

RuleEd2

{<TrustedGraphEd> author <Ed>; log:semantics ?TGEd. ?TGEd log:includes

{?O type:NonProfit, officer:<George>}. <ruleEd1 author <Ed>}.

ruleEd1 assertions ?ARule1. ?Arule1 log:semantics ?Arule1S.?ARule1S log:includes

{{ <act1> type Donate, recipient <?O>. <Ed> intends <act1>}

=> {<at> type Authorization, rule <RuleEd2>, action <act1>, value "Denied"}}

Stated in words, RuleEd2 stipulates that if Ed's trusted graph contains an Organization for which George is an officer, and Ed intends to take an action of type "Donate" whose recipient is this Organization, then the request will be denied.

Next, we look at George's privacy rule. It is a privacy rule because it refers to an action over personal information about George.

RuleGeorge1

{G :is {?O type:NonProfit; officer:<George>. <TrustedGraphGeorge> author <George>; log:semantics ?TGG. ?TGG log:includes {?O officer:<George>. <ruleGeorge1> author <George>}.

ruleGeorge1 assertions ?AGeorge1. ?AGeorge1 log:semantics ?AGeorge1S. ?Ageorge1S log:includes

{ <act1> type Read; object ?G. ?A intends <act1>}

=> {<at> type Authorization; rule <RuleGeorge1>; action <act1>; value "Denied"}}

Stated in words, RuleGeoge1 says that if a read operation is intended over the KG containing the information that George is an officer of some organization, the authorization for that operation will be denied.

The stipulation expressed in the transparency legislation for non-profits to divulge its officers can be expressed as

KG

NonProfitAct type Law.

RuleTransp

{<ruleTransp> author <Congress>. <ruleTransp> assertions ?ATransp. ?ATransp log:semantics ?ATranspS. ?ATranspS log:includes{?G is {?O type Nonprofit; officer ?Ofr.

<act1> type Read; object ?G. ?A intends <act1>}

=> {<at> type Authorization; rule <ruleTtransp>; action <act1>; value "Allowed"}}.

<ruleTransp> provenance ?PTransp. ?PTransp log:semantics ?PTranspS. ?PTranspS log:includes {<ruleTransp> prov:HasPrimarySource <NonProfitAct>}

Stated in words, RuleTransp (which assumes there is a statement affirming that NonProfitAct is a Law) says that any operation to read who are the officers of an organization of type NonProfit will be allowed. Furthermore, it also states that the provenance graph of RuleTransp contains a reference to the proper legal document, the text of the law (<NonProfitAct>) that is being interpreted by this rule.

In order to manage the conflict between RuleGeorge1 and RuleTransp, there is a meta-rule stipulating that the latter has precedence over the former. This precedence relation is used in the Sort-by-Precedence function call in line 5 in Fig. 4.

KG

PersonalPrivacyRule subClassOf Rule. Legislation subClassOf Rule.

MetaRule1

{<MetaRule1> assertions ?AMR1. ?AMR1 log:semantics ?AMR1S. ?AMR1S log:includes

{{?R1 type PersonalPrivacyRule. ?R2 type Legislation} => {?R2 precedes ?R1}}

Stated in words, this rule says that rules of type Legislation have precendence over rules of type PersonalPrivacyRule.

We must also define more precisely what is a PersonalPrivacyRule – it is any rule that uses a PersonalInformation property in its antecedent, which must also be defined.

?RuleP assertions ?RulePA.?RulePA log:semantics ?RulePAS;?RulePAS log:includes {{?p1 ?r ?p2.

(?p1 rdf:type Person OR ?p2 rdf:type Person)}}

=> {?r rdf:type PersonalInformationRelation}

Stated in words, this rule says that any relation involving a person is a PersonalInformationRelation. This is admittedly a simplification, as in actual situations these relations that characterize "personal information" should be further elaborated.

ISSN: 1839-4183

{?RulePP assertions ?RulePAS. ?RulePPAS log:semantics ?RulePPASS;

?RulePPAS log:includes{?RulePP antecedent ?RulePPAA. ?RulePPAA log:semantics ?RulePPAAS

?RulePPAAS log:includes {{?p ?r ?q.(?r rdf:type PersonalInformationRelation)}

=> {?RulePP type PersonalPrivacyRule}}

Stated in words, this rule says that a Personal Privacy Rule is a rule that involves (uses) a Personal Information Relation in its antecedent. Notice that this is also a meta-rule, since its objects are rule themselves.

The second scenario does not require any additional rules. It simply results in an "allowed" authorization, because no conflict arises between RuleGeorge1 and RuleTransp when applied to <*AuditInc*>, since it is not of type NonProfit, and therefore RuleTransp is not applicable. This illustrates a possible loophole in the legislation which could be avoided if the NonProfit legislation prohibited an organization from being audited by another organization having common officers.

7. CONCLUSIONS

We have presented a usage framework explicating the various types of specifications that must be made to capture privacy, transparency and trust concerns. The framework also provides a better understanding of the relations between these concerns.

Trust entails determining which data items will be used to perform an action; privacy and transparency involve controlling who can perform an action over a data item. Trust is thus more fundamental, as privacy and transparency rules must be based on trusted data. Furthermore, privacy and transparency are approached as being different points along the control dimension, thus explaining the natural tension between the two.

We have shown how legal requirements, and other types of norms, which ultimately regulate the functioning of any application that uses the KG, can be incorporated into the KG itself. The various nuances and interdependencies of these concerns were illustrated in a running example. One interesting point in the example is the fact that in spite of careful policies, loopholes in the regulations could allow undesired actions to take place.

As ongoing and future work, we are investigating implementation architectures to allow efficient and scalable usage control over existing KGs, as well as exploring the applicability to various domains. This architecture can also be fitted into wider legal governance models that have been developed in the literature in the last five years and into more general Online Dispute Resolution governance models (Ebner and Zeleznikow 2016). While the algorithm and the example presented here consider that the authorization for an action would be determined in an automated fashion, it is entirely possible a version where the algorithm could reach a state in which it is not able to resolve conflicts between rules. At this stage, it could refer the information and their supporting facts to a human being, to allow them to bring in additional criteria which would be outside the scope of the rules in the KG to reach a final decision. We are investigating the extension of the framework to accommodate this mode.

8. REFERENCES

1. A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A. N. Ngomo, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, and A. Zimmermann. "Knowledge Graphs". *arXiv*. abs/2003.02320. 2020.
2. Almendra, V. D. S., and Schwabe, D. 2006. Trust policies for semantic web repositories. In *Proceedings of 2nd International Semantic Web Policy Workshop* (SWPW'06), at the 5th International Semantic Web Conference, ISWC, pp. 17-31.
3. Angella J., Kim, K. and Johnson, K.P. 2016. "Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook", *Computers in Human Behavior,* 58: 98-108. http://dx.doi.org/10.1016/j.chb.2015.12.047 .
4. Artz, D., Gil, Y. 2007. "A survey of trust in computer science and the Semantic Web", *Web Semantics: Science, Services and Agents on the World Wide Web,* 5 82): 58-71. http://www.sciencedirect.com/science/article/pii/S1570826807000133
5. Australian Government. 2019. Department of the Treasury. *Regulating in the digital age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry*

(Government Response, December). https://treasury.gov.au/publication/p2019-41708

6. Australian Government. 2020. Attorney-General's Department. *Privacy Act Review. Issues Paper.* October, https://apo.org.au/sites/default/files/resource-files/2020-10/apo-nid309253.pdf
7. Berners-Lee, T., Connolly, D., Kagal, L., Scharf, Y., Hendler, J. 2008. "N3Logic: A Logical Framework for the World Wide Web", *Theory Pract. Log. Program.* 8: 249–269.
8. Bernstein, E. 2017. "Making Transparency Transparent: The Evolution of Observation in Management Theory", *Academy of Management Annals* 11 (1): 217 – 66.
9. Bizer, C. and Cyganiak, R. 2009. "Quality-Driven Information Filtering Using the WIQA Policy Framework", *Web Semantics: Science, Services and Agents on the World Wide Web*. 7: 1-10. http://www.websemanticsjournal.org/index.php/ps/article/view/157/155
10. Blasi-Casagran, C. 2014. "The EU Role in Shaping Global Data Privacy Standards", PrivOn2015 https://sites.google.com/site/privon2015/accepted-papers
11. Blasi-Casagran, C. 2016. *Global data protection in the field of law enforcement: an EU perspective.* London: Routledge.
12. Cameron. K. 2004. December- "The Seven laws of the Internet". http://www.identityblog.com/stories/2004/12/09/thelaws.html
13. Carlson, A., Betteridge, J., Wang, R.C., Hruschka Jr., E.R., Mitchell, T.M. 2010. "Coupled semi-supervised learning for information extraction". In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining,* pp. 101–110.
14. Casanovas, P., Arraiza, J., Melero, F., González-Conejero, J., Molcho, G. and Cuadros, M. 2014. "Fighting Organized Crime Through Open Source Intelligence: Regulatory Strategies of the CAPER Project". In Hoekstra, R. ed., *Legal Knowledge and Information Systems*: *JURIX 2014.* Amsterdam: IOS Press, pp. 189-198.
15. Casanovas, P., Jorge González-Conejero, J., de Koker, L. 2017. "Legal compliance by design (LCbD) and through design (LCtD): preliminary survey." TERECOM@JURIX 2017,
16. Casanovas, P., Palmirani, M., Peroni, S., Van Engers, T. and Vitali, F. 2016. "Semantic web for the legal domain: the next step". *Semantic Web, 7*(3): 213-227.
17. Castelfranchi, C., Falcone, R. Social Trust: A Cognitive Approach. In: Castelfranchi, C., Yao-Hua Tan (Eds.) *Trust and Deception in Virtual Societies.* Dartmouth: Springer-Verlag (2001).
18. Castells, M. 2010. *The Rise of the network society* (2 ed.). Cambridge, MA, USA: Blackwell Publ. Inc.
19. Cavoukian, A. 2006. "7 Laws of Identity: The Case for Privacy-Embedded Laws of Identity in the Digital Age", *Technology*, Ontario Information and Privacy Commissioner, October, pp. 1-24.
20. Cavoukian, A. 2010. "Privacy by Design. *The 7 Foundational Principles. Implementation and Mapping of Fair information Practices.* Information and Privacy Commissioner, Ontario, Canada.
21. Colesky, M., Hoepman, J.H. and Hillen, C. *2016, "A critical analysis of privacy design strategies". In 2016 IEEE Security and Privacy Workshops (SPW)*, IEEE, pp. 33-40.
22. Council of the European Union, European Parliament, 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council. General Data Protection Regulation"*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 .
23. Davenport, T.H. and Beck, J.C. 2001. *The attention economy: Understanding the new currency of business.* Harvard Business Press.
24. Ebner, N. and Zeleznikow, J. 2016. "No sheriff in town: governance for online dispute resolution", *Negotiation Journal, 32*(4): 297-323.
25. Ehrlinger, L. and Wöß, W. 2016. "Towards a Definition of Knowledge Graphs", SEMANTiCS, CEUR 1695, http://ceur-ws.org/Vol-1695/paper4.pdf
26. Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., and Vrandečic, D. 2014. "Introducing Wikidata to the Linked Data Web", Proc. ISWC 2014, Part I. LNCS, vol. 8796. Cham: Springer.
27. European Union. 2012. *Charter of Fundamental Rights of The European Union.* (2012/C 326/02) 26.10.2012 Official Journal of the European Union C 326/391. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT
28. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H, Metayer, D.L., Tirtea, R. and Schiffner, S. 2014. *Privacy and Data Protection by Design – From policy to engineering.* Technical report, ENISA, December. ISBN 978-92-9204-108-3, DOI 10.2824/38623. https://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-and-data-protection-by-design
29. Gandon, F. 2018. "A Survey of the First 20 Years of Research on Semantic Web and Linked Data", *Revue des Sciences et Technologies de l'Information* - Série ISI : Ingénierie des Systèmes d'Information, Lavoisier, https://hal.inria.fr/hal-01935898/document

ISSN: 1839-4183

30. Gerck, E. 1998. *Toward Real-World Models of Trust: Reliance on Received Information.* Report MCWG-Jan22-1998, http://www.safevote.com/papers/trustdef.htm
31. Golden, P. and Shaw, R., 2016. "Nanopublication beyond the sciences: the PeriodO period gazetteer". *PeerJ Computer Science, 2,* p.e44. https://peerj.com/articles/cs-44/
32. González-Fuster, G. 2014. *The emergence of personal data protection as a fundamental right of the EU,* LGT Vol. 16. Cham: Springer Science & Business.
33. Grandison, T., and Sloman, M. 2000. "A survey of trust in internet applications", *IEEE Communications Surveys & Tutorials* 3 (4): 2-16.
34. Groth, P., Gibson, A., and Velterop. J. 2010. "The anatomy of a nanopublication." *Information Services & Use* 30 (1-2): 51-56.
35. Hashmi, M., Casanovas, P. and de Koker, L. 2018. "Legal Compliance Through Design: Preliminary Results of a Literature Survey". TERECOM2018@JURIX, Technologies for Regulatory Compliance http://ceur-ws.org/Vol-2309/06.pdf
36. Hashmi, M., Governatori, G., Lam, H.P. and Wynn, M.T. 2018. "Are we done with business process compliance: state of the art and challenges ahead", *Knowledge and Information Systems, 57*(1): 79-133.
37. Heimstäd, M., Dobusch, L. 2018. "Politics of Disclosure: Organizational Transparency as Multiactor Negotiation", *Public Administration Review* 78 (5): 727-738
38. Hoepman, J.-H. 2018. "Privacy Design Strategies" (The Little Blue Book). https://repository.ubn.ru.nl/bitstream/handle/2066/195397/195397.pdf?sequence=1
39. Hoepman, J.-H. 2014. "Privacy Design Strategies (extended abstract)", in N. Cuppens-Boulahia et al. (Eds.), *ICT Systems Security and Privacy Protection* - 29th IFIP TC 11 International Conference, SEC 2014, Marrakech, Morocco, June 2-4, 2014. Proceedings. Springer 2014 *IFIP Advances in Information and Communication Technology*, pp. 446-459.
40. Kejriwal, M. 2019. *Domain-Specific Knowledge Graph Construction.* Cham: Brief Springer International Publishing
41. Koops, B.J. and Leenes, R. 2014. "Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design'provision in data-protection law". *International Review of Law, Computers & Technology,* 28(2): 159-171.
42. Kuhn, T., Meroño-Peñuela, A., Malic, A., Poelen, J.H., Hurlbert, A.H., Centeno, E., Furlong, L.I., Queralt-Rosinach, N., Chichester, C., Banda, J.M., Willighagen, E.L., Ehrhart, F., Evelo, C.T., Malas, T.B., & Dumontier, M. 2018. Nanopublications: A Growing Resource of Provenance-Centric Scientific Linked Data. CoRR, abs/1809.06532. https://arxiv.org/pdf/1809.06532.pdf
43. Langheinrich, M. 2001. "Privacy by design—principles of privacy-aware ubiquitous systems. In G.D. Abowd, B. Brumitt, and S. Shafer (Eds.), *International conference on Ubiquitous Computing,* LNCS 2201, Berlin, Heidelberg: Springer, pp. 273-291.
44. Laufer, C, Schwabe, D. 2017." On Modeling Political Systems to Support the Trust Process, Proceedings of the Privacy Online", PrivON 2017, co-located with the 16th International Semantic Web Conference (ISWC 2017) Vienna, Austria, Out. 2017, p.1 – 16 http://ceur-ws.org/Vol-1951/PrivOn2017_paper_7.pdf
45. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C.. 2013. "DBpedia-A large-scale, multilingual knowledge base extracted from Wikipedia", *Semantic Web Journal* 6(2): 167–195.
46. Lenat, D.B. 1995. "CYC: a large-scale investment in knowledge infrastructure", *Communications ACM* 38(11): 33–38.
47. Martín-Chozas, P., Montiel-Ponsoda, E. and Rodríguez-Doncel, V. 2019. "Language resources as linked data for the legal domain" In, G. Peruginelli, S. Faro (Eds), *Knowledge of the Law in the Big Data Age,* Amsterdam: IOS Press, pp. 170-180.
48. Marwick, A. andLewis, R.. 2017. "Media Manipulation and Disinformation Online." Data & Society Research Institute, New York. https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf
49. McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P. and Cimiano, P. 2017. "The Ontolex-Lemon model: development and applications". In *Proceedings of eLex 2017 Conference*, pp. 19-21.
50. Meijer, A. 2013. "Understanding the Complex Dynamics of Transparency", *Public Administration Review* 73 (3): 429 – 439.
51. Miller, G. A. 1995. "WordNet: Smith, H.J., Dinev, T., Xu, H A lexical database for English", *Communications ACM* 38 (11): 39–41.
52. Norris, C., De Hert, P., L'Hoiry, X., Galetta, A. (Eds.) 2017. *The Unaccountable State of Surveillance*. Cham: Springer.
53. Paci, F., Squicciarini, A., Zannone, N. 2018. "Survey on Access Control for Community-Centered Collaborative Systems". *ACM Comput. Surv*. 51 (6): 1–6.
54. Pagallo, U., Aurucci, P., Casanovas, P., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Schafer, B. and Valcke, P., 2019. *AI4People-On Good AI Governance: 14 Priority*

Actions, a SMART Model of Governance, and a Regulatory Toolbox. https://www.eismd.eu/pdf/AI4PEOPLE%20On%20Good%20Ai%20Governance%202019.pdf

55. Pagallo, U., Casanovas, P. and Madelin, R. 2019. "The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data". *The Theory and Practice of Legislation, 7* (1): 1-25. https://www.tandfonline.com/doi/full/10.1080/20508840.2019.1664543
56. Peczenik, A. 2004. *Scientia Juris.* In *A Treatise of Legal Philosophy and General Jurisprudence,* Vol. 4. Dartmouth: Springer.
57. Pfitzmann, A. and Hansen, M. 2010. "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management". V.034. http://www.maroki.de/pub/dphistory/2010_Anon_Terminology_v0.34.pdf
58. Pinyol, I.; Sabater-Mir, J.; Computational trust and reputation models for open multi-agent systems: a review. Artificial Intelligence Review 40:1–25 DOI 10.1007/s10462-011-9277-z (2013)
59. Poblet, M., Casanovas, P. and Rodríguez-Doncel, V. 2019. *Linked Democracy: Foundations, tools, and applications.* Springer Law Brief n. 750. Cham: Springer Nature.
60. Rastogi, N., Gloria, M.J.K. and Hendler, J. 2017. "Security and Privacy of performing Data Analytics in the cloud-A three-way handshake of Technology, Policy, and Management". *arXiv preprint arXiv:1701.06828.*
61. Rehm, G., Galanis, D., Labropoulou, P., Piperidis, S., Welß, M., Usbeck, R., Köhler, J., Deligiannis, M., Gkirtzou, K., Fischer, J. and Chiarcos, C., 2020. "Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability". *arXiv preprint arXiv:2004.08355.*
62. Rifkin, J. (2011). The third industrial revolution: How lateral power is transforming energy, the economy, and the world. New York: Palgrave Macmillan.
63. Rodriguez-Doncel, V., Santos, C., Casanovas, P. and Gomez-Perez, A. (2016). "Legal aspects of linked data–The European framework". *Computer Law & Security Review, 32* (6): 799-813.
64. Rodríguez-Doncel, V., Santos, C., Casanovas, P. and Gómez-Pérez, A. (2015) "A Linked Term Bank of Copyright-Related Terms". In *JURIX-2015*, Amsterdam: IOS Press, pp. 91-100.
65. Sartor, G. 2005. *Legal reasoning.* In A *Treatise of Legal Philosophy and General Jurisprudence.* Vol. 5. Dartmouth: Springer.
66. Schudson, M. 2015. *The Rise of the Right to Know: Politics and the Culture of Transparency,* 1945-1973. Cambridge, MA: Harvard University Press.
67. Schwab, K., 2017. *The fourth industrial revolution.* Crown Business.
68. Schwabe, D., Laufer, C., and Busson, A. 2019. "Building Knowledge Graphs About Political Agents in the Age of Misinformation, http://arxiv.org/abs/1901.11408
69. Sherchan, W. Nepal, S., and Paris, C. 2013. "A Survey of trust in social networks", *ACM Comput. Surv.* 45, 4, Article 47 , August.
70. Smith, H.J., Dinev, T., Xu, H. 2011. "Information Privacy Research: An Interdisciplinary Review", *MIS Quarterly* 35: 989–1015.
71. Speer, R. and Havasi, C. 2012. "Representing General Relational Knowledge" In *ConceptNet 5, Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'12), ELRA, pp. 3679-3686.
72. Such, J.M. and Criado, N. 2018. "Multiparty Privacy in Social Media", *Communications ACM.* 61: 74–81.
73. Suchanek, F.M., Kasneci, G., and Weikum, G.. 2007. "YAGO: a core of semantic knowledge unifying WordNet and Wikipedia". In *16th International Conference on World Wide Web,* pp. 697–706.
74. Tavani, H.T. and Moor, J.H. 2001. "Privacy Protection, Control of Information, and Privacy-enhancing Technologies", *SIGCAS Comput. Soc*. 31: 6–11.
75. Tavani, H.T. 2007. "Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy", *Metaphilosophy* 38: 1–22.
76. Torra, V. 2017. *Data Privacy: Foundations, New Developments and the Big Data Challenge.* Cham: Springer Nature.
77. United Nations, Universal Bill of Human Rights, Resolution A/RES/217(III) http://unbisnet.un.org:8080/ipac20/ipac.jsp?session=14O243550E15G.60956&profile=voting&uri=full=3100023~!909326~!67
78. US Privacy Protection Study Commission. 1977. *Personal Privacy in an Information Society.* https://epic.org/privacy/ppsc1977report/
79. US Secretary's Advisory Committee on Automated Personal Data Systems. 1973. *Records, Computers and the Rights of Citizens.* https://aspe.hhs.gov/report/records-computers-and-rights-citizens
80. Vrandečić, D., Krötzsch, M. 2014. "Wikidata: a free collaborative knowledge base", *Communications ACM* 57 (10): 78–85.

ISSN: 1839-4183

81. Westin, A. 1967. *Privacy and Freedom*. New York: Atheneum.
82. Zou, L., and Özsu, M.T. 2017. "Graph-based RDF data management", *Data Science and Engineering* 2 (1): 56-70.
83. Zuboff, S., 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Barack Obama's Books of 2019. NY: Profile Books.