# DALICC: A LICENSE MANAGEMENT FRAMEWORK FOR DIGITAL ASSETS

Tassilo Pellegrini<sup>1</sup>, Giray Havur<sup>2</sup>, Simon Steyskal<sup>2</sup>, Oleksandra Panasiuk<sup>3</sup>, Anna Fensel<sup>3</sup>, Victor Mireles<sup>4</sup>, Thomas Thurner<sup>4</sup>, Axel Polleres<sup>2</sup>, Sabrina Kirrane<sup>2</sup>, Andrea Schönhofer<sup>1</sup>

<sup>2</sup> Vienna University of Economics and Business, Department of Informations Systems & Operations, Welthandelsplatz 1, 1020 Wien, AT, [firstname.lastname]@wu.ac.at, https://www.wu.ac.at/en/infobiz/

Keywords: Derivative Works, License Clearance, Legal Technologies, Reasoning

Abstract:

This paper describes the Data Licenses Clearance Center, a software framework that supports the cost-efficient and transparent resolution of licensing conflicts that occur in the reutilization of digital assets. DALICC provides a library of machine readable standard licenses and allows users to compose arbitrary custom licenses. In addition, the system supports the clearance of rights issues by providing users with information about the equivalence, similarity and compatibility of licenses. A public beta version of the system is available at <a href="https://www.dalicc.net/">https://www.dalicc.net/</a>.

### 1. Introduction

DALICC stands for Data Licenses Clearance Center. It is a software framework that supports legal experts, innovation managers and application developers in the legally secure reutilization of third

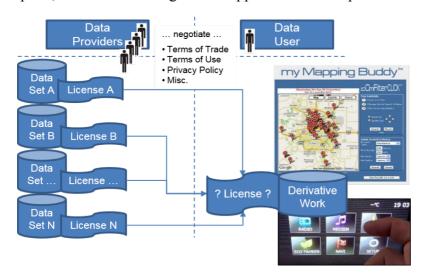


Figure 1: Licenses clearance in derivative works

party data sources. The DALICC framework enables automated clearance of rights, thus helping to detect licensing conflicts and significantly reducing the costs of rights clearance in the creation derivative works. This is insofar necessary as modern applications increasingly retrieve, store and process data from a variety of sources. This can raise questions about the compatibility of licenses and the application's compliance with existing law. In order to provide commercial products

<sup>&</sup>lt;sup>1</sup> Fachhochschule St. Pölten, Department Medienwirtschaft, Matthias Corvinus Straße 15, 3100 St. Pölten, AT, tassilo.pellegrini@fhstp.ac.at; <a href="http://www.fhstp.ac.at">http://www.fhstp.ac.at</a>

<sup>&</sup>lt;sup>3</sup> STI Innsbruck, University of Innsbruck, Technikerstraße 21a, 6020 Innsbruck, AT, [firstname.lastname]@sti2.at, https://www.sti-innsbruck.at/

<sup>&</sup>lt;sup>4</sup> The Semantic Web Company, Neubaugasse 1, 1070 Vienna, AT, [firstname.lastname]@semantic-web.com, www.semantic-web.at

services on top of third party data, license clearance is necessary to assure legal compatibility (Hoffmann et al. 2015).

Publishing data and reusing it for commercial or non-commercial purposes as depicted in Figure 1 has become a common practice and cornerstone of the so called digital economy (World Bank 2014). The growing popularity especially of protective and permissive licenses (some rights reserved) has added to the complexity of rights clearance in the commercial exploitation of derivative works. As a consequence a wide array of data publishing guidelines were recommended (Guibault 2011; Hyland & Wood 2011; Frosterus et al. 2011) giving expression to the fact that licensing of data is a fairly new kind of economic practice and still subject to debate concerning the adequate design and management of licensing policies (Sonntag 2006; Sande et al. 2012; Archer 2013; Pellegrini 2014; Ermilov & Pellegrini 2015).

New data practices stimulated by phenomena like open data, open innovation, and crowdsourcing initiatives as well as the increasing interconnection of services, sensors, and (cyberphysical) systems have nurtured an environment, in which the effective handling of licenses has become key to innovation, productivity and value creation. According to the OECD the effective management of intangible assets is the primary driver of innovation in the ICT-enabled service sector and source of competitive advantage at the macro- and micro-level (OECD 2008). This line of argument corresponds with a study conducted by Oxford Economics that argue that "insights derived by linking previously disparate bits of data can become the sparks that ignite rapid innovation" (Roehring & Pring 2013). But according to the EU Agency for Network and Information Security the main obstacle in the digital ecosystems of the future is the legal impact of information exchange (ENISA 2013). This is especially relevant in the context of the European strategy for a data-driven economy which aims to "nurture a coherent European data ecosystem, stimulate research and innovation around data and improve the framework conditions for extracting value out of data" (European Commission 2014).

But clearing and negotiating rights issues is a time-consuming, complex and error-prone task. Challenges associated with clearance issues are (1) high transaction costs in the manual clearance of licensing terms and conditions, (2) sufficient expertise to detect compatibility conflicts between two or more licenses, and (3) negotiation and resolution of licensing conflicts between involved parties.

To tackle the problems mentioned above, the DALICC software framework will develop and integrate various functionalities that allow the automated clearance of rights issues. To do so, the following technical problem areas will be addressed.

## 2. State of the Art in Automated Rights Clearance

Licenses express deontic statements (permissions, prohibitions and obligations) associated with a protectable asset as defined by copyright law or contract law. Licenses control access to, usage of, and transactions on top of digital assets, be it under conditions of property rights (all rights reserved) or public domain (no rights reserved) (Guibault 2011). Figure 2 depicts the spectrum of available licensing models available to the management of digital assets.



Figure 2: Spectrum of licensing models

Rights Expression Languages (RELs) are a subset of Digital Rights Management technologies that are used to explicate machine-readable deontic statements for purposes of Digital Asset Management (Pellegrini et al. 2018). Recent research conducted on the genealogy of RELs indicate

that since 1989 more than 60 RELs have been developed from which 23 can be used to express licenses (Pellegrini et al. 2018a). Among the most prominent REL vocabularies used to represent licenses are MPEG-21<sup>1</sup>, the Creative Commons Rights Expression Language (ccREL)<sup>2</sup> and the Open Digital Rights Language (ODRL)<sup>3</sup>. When it comes to machine-processing of licensing information various approaches exist that address this problem. An early proposal for a generic logic for reasoning over licenses is provided by Pucella & Weissman (2002), but it has not been implemented with existing RELs like ODRL or MPEG-21 nor has it been evaluated in practice. Garcia et al. (2009/2007/2004) propose an OWL ontology to describe copyright issues in closed datasets for rights clearance purposes. Their approach is based on a deprecated version of the ODRL vocabulary and constitutes a proof of concept that has not been implemented and tested against issues arising from contemporary open data licensing. More recent work is provided by Hosking et al. (2014) who present a rule-based engine, built on top of the Carneades Framework (Gordon et al. 2007), to reason over various sets of licenses, while additionally suggesting potential licenses by which to safely share derived outputs. Instead of applying deductive reasoning they used a non-monotonic formalism suitable for modelling situations in which contradictory statements are being processed. Villata and Gandon (2012) and Governatori et al. (2013) describe the formalization of a license composition tool for derivative works. They extend their research by introducing semantics based on a deontic logic (Rotolo et al. 2013; Governatori et al. 2014; Rodríguez-Doncel et al. 2015) for the comparison of the permissions, prohibitions and duties stated in a given license. They also provide a demo called Licentia<sup>4</sup> that exemplifies the practical value of such a service. This line of work is an interesting approach to detect and potentially solve licensing conflicts e.g. by composing a new license that resolves the conflict. The pitfall of their approach is that in the current version license compatibility can just be checked against a bundle of selected permissions, obligations and prohibitions and not against a selection of two or more other licenses containing these conditions. Additionally, their compatibility check assumes a reciprocal relationship between licenses instead of a directed relationship as given under real-world circumstances.

# 3. DALICC Software Architecture & Functional Spectrum

The DALICC framework consists of four main functional components, namely: license library, license search, license composer, and license substitutability check, as shown in Figure 3. Technology-wise, the DALICC framework utilizes the following components: a Virtuoso triplestore<sup>5</sup>, a Drupal<sup>6</sup> based web application, the PoolParty Semantic Suite<sup>7</sup>, and a Clingo Answer Set Programming (ASP) reasoner (Gebser et al. 2014).

\_

The Moving Pictures Experts Group, 2019. MPEG-21 | MPEG [WWW Document]. URL https://mpeg.chiariglione.org/standards/mpeg-21 (accessed 1.6.19).

<sup>&</sup>lt;sup>2</sup> W3C, 2008. ccREL: The Creative Commons Rights Expression Language [WWW Document]. URL https://www.w3.org/Submission/ccREL/ (accessed 1.6.19).

<sup>&</sup>lt;sup>3</sup> W3C, 2016. ODRL Version 2.2 Ontology [WWW Document]. URL https://www.w3.org/ns/odrl/2/ (accessed 1.6.19).

<sup>&</sup>lt;sup>4</sup> INRIA, 2014. Licentia by INRIA [WWW Document]. URL http://licentia.inria.fr/ (accessed 1.6.19).

<sup>&</sup>lt;sup>5</sup> OpenLink Software, 2019. Virtuoso RDF Triple Store [WWW Document]. URL https://virtuoso.openlinksw.com/rdf/ (accessed 1.6.19).

<sup>&</sup>lt;sup>6</sup> Drupal, 2019. Drupal - Open Source CMS | Drupal.org [WWW Document]. URL https://www.drupal.org/ (accessed 1.6.19).

<sup>&</sup>lt;sup>7</sup> Semantic Web Company, 2019. PoolParty Semantic Suite - Your Complete Semantic Platform [WWW Document]. URL https://www.poolparty.biz/ (accessed 1.6.19).

The license library (Figure 4) is a repository that contains machine-readable and human-readable representations of licenses, the former as ODRL policies, and the latter as plain text. Licenses properties are queried using SPARQL and presented to the user in an easily digestible manner.

The license search (Figure 5) allows to search for licenses either by full-text or by selecting specific actions (permissions, duties or obligations) from a questionnaire. For example, when the action commercial use is set to "yes", the corresponding SPARQL query reveals that this question is linked, via the dalicc:needsPermission predicate to cc:CommercialUse. The cc:CommercialUse permission is then translated into a rule and processed by a reasoner. Details about the reasoning engine are described below.

The license composer (Figure 6) allows to create customized licenses from a set of questions which are mapped to ODRL, ccREL and the DALICC vocabularies. The user must declare various provenance information about the asset and then can specify the licensing terms by defining permissions, duties and prohibitions. After finishing the editing process the custom license can be downloaded as machine-readable RDF file or human-readable text-file.

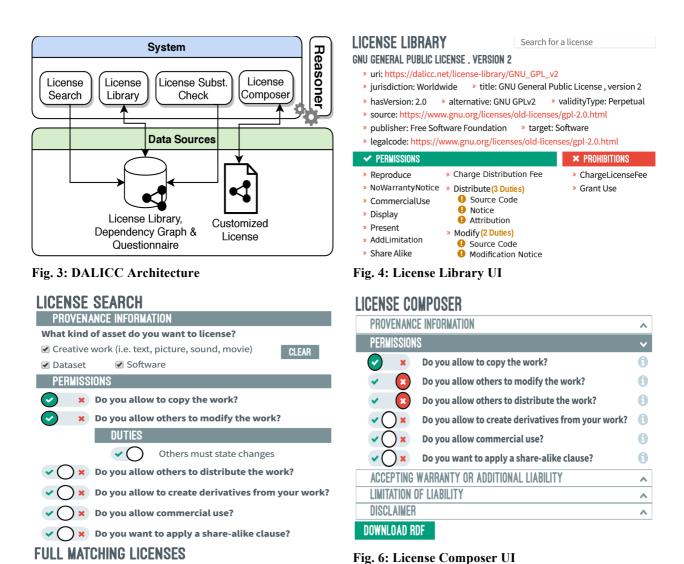


Fig. 5: License Search UI

» Attribution 4.0 International» GNU Lesser General Public License

When it is necessary to license a derivative work consisting of components that come with various licenses (i.e., initial licenses), each license must be substitutable with the target license, under which

» The MIT License

the derivative shall be made available. For instance, it is allowed to use an asset licensed under APACHEv2 in a GPLv3 project, but not vice versa. DALICC allows to perform such checks and indicates conflicts at the action level of each license.

## 4. Data Modelling

To gain a valid, machine-readable representation of a license, our modelling process consists of four phases: (i) analyzing the license text; (ii) defining vocabularies to express the license's terms; (iii) deriving modelling and mapping mechanisms, and (iv) defining a common model for annotating and comparing licenses. For the analytic part, we selected 54 commonly used licenses which cover the spectrum of digital assets relevant to DALICC, namely datasets, software and creative work. We subsequently went through the legal text of each license and identified permissions, associated duties (obligations) and prohibitions. In order to represent license concepts in a structured machine-readable format we chose the ODRL information model for modelling licenses in the form of policies that express permissions, prohibitions and duties related to the usage of assets. ODRL also defines a vocabulary of general terms (e.g., odrl:reproduce, odrl:distribute, odrl:modify) and can be further extended with terms from other vocabularies such as CC REL (e.g., cc:CommercialUse, cc:DerivativeWorks)<sup>10</sup>. To express legally valid machine-readable representations of licenses, it was necessary to create additional terms (e.g., dalicc:perpetual as a validity type, dalicc:worldwide as a jurisdictional property, dalicc:chargeDistributionFee as permission and prohibition actions, and dalicc:modificationNotice as a duty action), which we called the DALICC vocabulary extension.

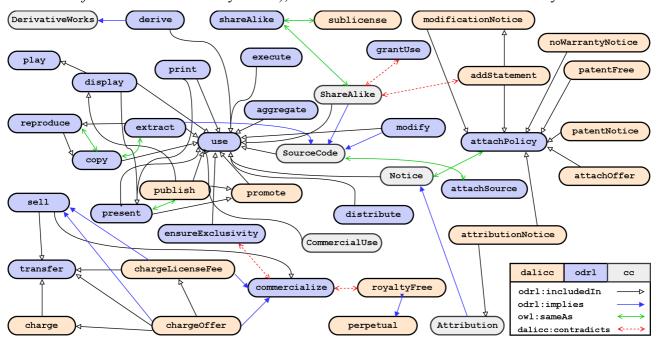


Fig. 7: DALICC Dependency Graph

To express the explicit and implicit relationships defined between actions, we use a dependency graph (Figure 7) following the work of Steyskal & Polleres (2015). It represents hierarchical relationships (e.g., odrl:distribute odrl:includedIn odrl:use), implications derived from a specific action (e.g., dalicc:modificationNotice odrl:implies odrl:modify), equalities (e.g., odrl:reproduce owl:sameAs cc:Reproduction), and contradictions (e.g., cc:ShareAlike dalicc:contradicts odrl:grantUse). The dependency graph semantically connects the legal domain graph on the one side to the schematic policy representations on the other. Additionally, it is an important mediating layer for the execution of reasoning and inferencing mechanisms described in the next section.

## 5. Reasoning over Licenses

To reason over licenses we use Answer Set Programming (ASP) (Brewka et al. 2011), a declarative (logic-programming-style) paradigm for solving combinatorial search problems by defining and evaluating rule sets. Sets of rules are evaluated in ASP under the so-called stable-model semantics, which allows several models (i.e., answer sets). We use the state-of-the-art ASP system CLINGO (Gebser et al. 2014), as this ASP system is among the most efficient implementations (Gebser et al. 2017).

Conflict detection: This component checks the logical coherence of the created license, provides information on equivalence, similarity and compatibility with other licenses, and is to support our work in conflict resolution between licenses. Policies should be understood as a set of rules derived from the RDF<sup>8</sup> graphs of the licenses. Herein, a rule that permits or prohibits the execution of an action on certain assets does not only affect other rules that govern the execution of the same action on the same asset(s) but also those permitting or prohibiting related actions on the same asset(s). In this sense, CLINGO is not only an alternative to extensive materialization, which in this case is essential for search, but also enables listing sets of compatible statements. This latter possibility is necessary for effective computation of conflicts between licences, in particular for identifying the conflicting and non-conflicting parts of a license. The reasoning functionality is wrapped by a webservice that: (i) does the translation of RDF descriptions of licenses to CLINGO statements (ii) transforms user's queries into ASP queries (iii) parses the outputs and (iv) handles the composition of all statements to be passed to the reasoner (e.g. dependency graph, ODRL axioms, license rules).

**Identifying suitable licenses:** Another functionality provided by the reasoner is the identification of suitable licenses. To do so, we extended the ASP program to cover the detection of conflicting/non-conflicting licenses by re-utilizing the input selected in the UI. In the user-facing platform provided by the DALICC system, it is possible to search for licenses by answering a series of questions regarding permissions, prohibitions, and duties. When a user selects a combination of these, a query is issued to the reasoner, which returns all licenses from the library which are completely or partially non-conflicting with the given answers.

**Substitutability check:** Another issues critical to license clearance is to take the directionality of licensing into account. Hence, we check the substitutability of an initial license towards a target license by (1) comparing sets of permissions, prohibitions and duties between licenses, (2) checking if the initial and the target licenses are equal, i.e., if they have the same set of permissions, prohibitions and duties, (3) testing if the initial license has a *odrl:shareAlike* or *cc:ShareAlike* rules, and fails if there is such a rule and the licenses are not equal, and (4) inferencing if the initial license has a *odrl:nextPolicy* rule, and fails if *odrl:nextPolicy* is not the target license.

### 6. Assessment of the DALICC Framework

In order to ensure the legal validity of the machine readable licenses and the corresponding license compatibility assessment, both the development and the testing of the platform's components have been carried out under the auspices of an Austrian law firm<sup>9</sup> specialized in the subject matter.

At the time of writing, we carried out a usability pre-test to assess the ease of use and the comprehensiveness of the interface and interaction design and collect input for a major usability test scheduled for November 2018. We collected feedback from 12 users from which 6 users were software developers with basic knowledge on licensing, while the remaining 6 users were legal professionals in the domain. The system was evaluated using representative sample scenarios

<sup>&</sup>lt;sup>8</sup> W3C, 2014. RDF - Semantic Web Standards [WWW Document]. URL https://www.w3.org/RDF/ (accessed 1.6.19).

<sup>&</sup>lt;sup>9</sup> Hoehne, In der Maur & Partner Rechtsanwälte GmbH & Co KG. (https://https://h-i-p.at/)

("search for a specific license", "create a new license", "find equivalent licenses", "find licensing conflicts", "select compatible licenses"), which were used as a means to assess the general applicability of the developed artefacts and to provide feedback for further enhancement of the system. Most criticism related to the complexity of both the composer interface and the visualization of compatibility logs. The suggested improvements from the pre-test will be implemented in the next release of the DALICC framework planned for February 2019.

#### 7. Conclusion & Future Work

In this paper, we proposed a system that automatically supports license creation, license compatibility and license substitutability checking.

The potential for further work directions are as follows: (i) extending the license library with additional standard licenses as suggested by Ermilov & Pellegrini (2015) which would amount for approximately 150 standard licenses; (ii) utilizing already existing capabilities of the reasoning component for con

flict resolution. (iii) providing organizations with the means to create their own applications and workflows using DALICC APIs; (iv) visualizing data workflows taking into account the license provenance information; (v) providing license management schemes and security features hat advance the design and implementation of data and content value chains by rewarding both data and content providers that adhere to licensing obligations, possibly by utilizing blockchain protocols or other distributed ledger technologies.

## 8. Acknowledgements

The DALICC project (No. 855396) is a cooperative project funded by the Austrian Research Promotion Agency in the timeframe of 01.11.2016 to 30.03.2019. Details are available at www.dalicc.net.

#### 9. Literature

Archer, Phil et al. Study on business models for Linked Open Government Data. ISA programme by PwC EU Services. European Union, 2013.

Brewka, G., Eiter, T., Truszczyński, M., 2011. Answer set programming at a glance. Communications of the ACM 54, 92. https://doi.org/10.1145/2043174.2043195

ENISA (2013). Detect, SHARE, Protect Solutions for Improving Threat Data Exchange among CERTs, October 2013.

Ermilov, I., & Pellegrini, T. (2015). Data licensing on the cloud: empirical insights and implications for linked data (S. 153–156). ACM Press.

European Commission (2014). Towards a thriving data-driven economy. Brussels, 2.7.2014, COM(2014) 442 final.

Frosterus, M., Hyvönen, E., & Laitio, J. (2011). Creating and Publishing Semantic Metadata about Linked and Open Datasets. In D. Wood (Hrsg.), Linking Government Data (S. 95–112). New York, NY: Springer New York.

García, R., & Gil, R. (2009). Copyright Licenses Reasoning an OWL-DL Ontology. In Proceedings of the 2009 Conference on Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood (p. 145–162). Amsterdam: IOS Press.

García, R., Gil, R., & Delgado, J. (2007). A web ontologies framework for digital rights management. Artificial Intelligence and Law, 15(2), 137–154. http://doi.org/10.1007/s10506-007-9032-6.

García, R., Gil, R., & Delgado, J. (2004). Intellectual Property Rights Management Using a Semantic Web Information System. In R. Meersman & Z. Tari (Hrsg.), On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE (Bd. 3290, S. 689–704). Berlin, Heidelberg: Springer Berlin Heidelberg. Abgerufen von http://link.springer.com/10.1007/978-3-540-30468-5\_44.

Gebser, M., Kaminski, R., Kaufmann, B., Schaub, T., 2014. Clingo = ASP + Control: Preliminary Report. arXiv:1405.3694 [cs].

Gebser, M., Maratea, M., Ricca, F., 2017. The Sixth Answer Set Programming Competition. Journal of Artificial Intelligence Research 60, 41–95. https://doi.org/10.1613/jair.5373

Gordon, T.F., Prakken, H., Walton, D., 2007. The Carneades model of argument and burden of proof. Artificial Intelligence 171, 875–896. https://doi.org/10.1016/j.artint.2007.04.010

Governatori, G., Ho-Pun, L., Antonino, R., Serena, V., Fabien, G., 2013. Heuristics for Licenses Composition. Frontiers in Artificial Intelligence and Applications 77–86. https://doi.org/10.3233/978-1-61499-359-9-77

Governatori, G., Lam, H.-P., Rotolo, A., Villata, S., Auguste Atemezing, G., & Gandon, F. (2014). LIVE: a Tool for Checking Licenses Compatibility between Vocabularies and Data (Bd. 1272). Abgerufen von https://hal.inria.fr/hal-01076619.

Guibault, L. M. (2011). Open content licensing: from theory to practice. Amsterdam: Amsterdam Univ. Press.

Hosking, R., Gahegan, M., Dobbie, G., 2014. An eScience Tool for Understanding Copyright in Data Driven Sciences. IEEE, pp. 145–152. https://doi.org/10.1109/eScience.2014.37

Hoffmann, A., Schulz, T., Zirfas, J., Hoffmann, H., Roßnagel, A., & Leimeister, J. M. (2015). Legal Compatibility as a Characteristic of Sociotechnical Systems. Business & Information Systems Engineering, 57(2), 103–113. http://doi.org/10.1007/s12599-015-0373-5.

Hyland, B., & Wood, D. (2011). The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web. In D. Wood (Hrsg.), Linking Government Data (S. 3–26). New York, NY: Springer New York.

OECD (2008). Intellectual Assets and Value Creation. See also: http://www.oecd.org/sti/inno/40637101.pdf.

Roehring, Paul & Pring, Ben (2013). The Value of Signal and the Cost of Noise. London: Oxford Economics.

Pellegrini, T. (2014). Linked Data Licensing – Datenlizenzierung unter netzökonomischen Bedingungen. In E. Schweighöfer et al. (Hrsg.), Transparenz. 17. Int. Rechtsinformatik Symposium IRIS 2014. Wien: OCG Verlag.

Pellegrini, T., Mireles, V., Steyskal, S., Panasiuk, O., Fensel, A., Kirrane, S., 2018. Automated Rights Clearance Using Semantic Web Technologies: The DALICC Framework, in: Hoppe, T., Humm, B., Reibold, A. (Eds.), Semantic Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 203–218. https://doi.org/10.1007/978-3-662-55433-3 14

Pellegrini, T., Schönhofer, A., Kirrane, S., Steyskal, S., Fensel, A., Panasiuk, O., Mireles-Chavez, V., Thurner, T., Dörfler, M., Polleres, A., 2018a. A Genealogy and Classification of Rights Expression Languages – Preliminary Results, in: Data Protection / LegalTech - Proceedings of the 21st International Legal Informatics Symposium IRIS 2018, Colloquium. Presented at the IRIS 2018 - 21st International Legal Informatics Symposium, Editions Weblaw, Salzburg, Austria, pp. 243–250.

Pucella, R., & Weissman, V. (2002). A Logic for Reasoning about Digital Rights. In Proceedings of the 15th IEEE Workshop on Computer Security Foundations (p. 282–294). Washington, DC, USA: IEEE Computer Society.

Rodriguez, E., Delgado, J., Boch, L., & Rodriguez-Doncel, V. (2015). Media Contract Formalization Using a Standardized Contract Expression Language. IEEE MultiMedia, 22(2), 64–74. http://doi.org/10.1109/MMUL.2014.22.

Rotolo, A., Villata, S., Gandon, F., 2013. A Deontic Logic Semantics for Licenses Composition in the Web of Data, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, ICAIL '13. ACM, New York, NY, USA, pp. 111–120. https://doi.org/10.1145/2514601.2514614

Sande, Miel Vander; Portier, Marc; Mannens, Erik; Van de Walle, Rik (2012). Challenges for open data Usage: Open Derivatives and Licensing. In: https://www.w3.org/2012/06/pmod/pmod2012\_submission\_4.pdf, accessed February 12, 2016

Sonntag, Michael (2006). Rechtsschutz für Ontologien. In e-Staat und e-Wirtschaft aus rechtlicher Sicht. Stuttgart: Richard Boorberg Verlag.

Steyskal, S., Polleres, A., 2015. Towards Formal Semantics for ODRL Policies, in: Bassiliades, N., Gottlob, G., Sadri, F., Paschke, A., Roman, D. (Eds.), Rule Technologies: Foundations, Tools, and Applications. Springer International Publishing, Cham, pp. 360–375. https://doi.org/10.1007/978-3-319-21542-6\_23

Villata, S. and Gandon, F. (2012). Licenses compatibility and composition in the web of data. In: COLD - Workshop in conjunction with the 11th International Semantic Web Conference 2012. CEUR WS Proceedings, 905.

World Bank (2014). Open Data for Economic Growth. See also http://www.worldbank.org/content/dam/Worldbank/document/Open-Data-for-Economic-Growth.pdf.