

SUPPORTING AUTOMATED LICENSE CLEARANCE WITH THE DALICC FRAMEWORK

Tassilo Pellegrini

Fachhochschule St. Pölten, Department Medienwirtschaft
Matthias Corvinus Straße 15, 3100 St. Pölten, AT
Tassilo.pellegrini@fhstp.ac.at; <http://www.fhstp.ac.at>

Keywords: *Derivative Works, License Clearance, Legal Technologies, Reasoning, IRIS 2017*

Abstract: *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 party data sources. The DALICC framework enables the automated clearance of rights, thus helping to detect licensing conflicts and significantly reducing the costs of rights clearance in the creation of derivative works. This is insofar necessary as modern IT 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.*

1. Introduction

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]. In order to provide commercial products and services on top of third party data, license clearance is necessary to assure legal compatibility [HOFFMANN ET AL. 2015]. 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,
3. negotiation and resolution of licensing conflicts between involved parties.

To tackle the problems described above, the DALICC software framework will develop and integrate various functionalities that allow the automated clearance of rights issues.

2. The DALICC Framework

2.1. System Requirements & Affordances

The following requirements will be addressed the DALICC system:

Tackling license heterogeneity: In the creation of derivative works the simplest approach is to only combine content under the same well-known licence. This is over-restrictive though, and many licences under various names may permit their content to be combined. It is, however, difficult to judge, whether it is permitted and how the resultant content should be licensed. There may still be subtleties arising from unclear definitions of terms (i.e. «open» or «commercial use»), special clauses (i.e. share-alike) or implicit preconditions (i.e. «everything not permitted is forbidden» or «CC0 apart from images - see restrictions in further links»).

DALICC aims to resolve these issues by producing an audited set of machine-readable representations of licenses that allow licenses to be compared to each other in order to identify equivalent licenses and to point the user to potential conflicts if various licenses are being combined.

Tackling REL heterogeneity: Combining licences is simpler if all of the licences involved are expressed through the same REL. But over time various RELs (i.e. MPEG21, ODRL, ccREL) have emerged for various purposes, each providing their own vocabulary and level of expressivity. Hence, it is difficult to compare licenses that have been represented by different RELs. Additionally, it can sometimes be reasonable to extend the semantic expressivity of a given REL by adding expressions from another REL to cover the requirements of a real world scenario.

DALICC will solve this problem by applying a semantic web approach to represent RELs, map their terms to each other and extend their expressivity. It will represent existing vocabularies from various RELs by utilizing W3C-approved standards, thus allowing mappings between various RELs to be created. The output from this undertaking is the basis for the reasoning engine described in the next section.

Compatibility check, conflict detection & neutrality of the rules: A regular problem with semantic translation between schemas (such as RELs) is in making sure that the meaning of different terms are aligned. The regular problems are in demonstrating the equivalence of classes, properties and instances. To a certain degree this will be resolved by applying Semantic Web standards, but mapping alone won't solve the issue. More elaborated techniques like reasoning and inferencing mechanisms are necessary to improve the precision of conflict detection.

To solve these issues DALICC will create a reasoning engine that provides users with guided assistance in the detection of conflicts in accordance with specific usage scenarios of an envisioned application. In other words, DALICC will assist users with a workflow specifying the usage context, thus collecting supplementary information that helps to detect ambivalent concepts and potential conflicts. Based on this information the DALICC reasoning engine will reason over the set of licenses and infer instructions for the end-user on how to proceed in the license clearance process.

Legal validity of representations and machine recommendations: The semantic complexity of licensing issues means that the semantics of RELs must be clearly aligned within the specific application scenario. This includes a correct interpretation of the various national legislations according to the country of origin of a jurisdiction (i.e. German Urheberrecht vs. US copyright), the resolution of problems that are derived from multilinguality (i.e. multiple connotations of «royalties») and the consideration of existing case law in the resolution of licensing conflicts (i.e. Versata vs. Ameriprise)¹.

To tackle this, legal experts from inside and outside the DALICC consortium will check the legal validity of machine-readable licenses and the output of the reasoning engine for compatibility with applicable laws.

¹ See also: Versata, Trilogy Software, Inc. and Trilogy Development Group v. Ameriprise, Ameriprise Financial Services, Inc. and American Enterprise Investment Services, Inc., Case No. D-1-GN-12-003588; 53rd Judicial District Court of Travis County, Texas.

In several iteration cycles the DALICC output will be tested against laws and jurisdictions, checked for its semantic precision deriving from various languages and adjusted accordingly.

DALICC User-Interface and Interaction Design: Licenses in general and rights clearance in particular are complex topics that require a high level of problem awareness and legal expertise. Due to the abstractness and complexity of the topic especially non-legal professionals have to invest a lot of time and/or money to acquire this knowledge and search for viable solutions.

DALICC will test and provide easy-to-use user interfaces that help non-legal professionals to quickly learn about the issues of data licensing and understand the consequences of unsolved licensing conflicts of their application. Thus, DALICC will leverage the overall awareness and expertise on licensing issues, reduce the barriers of entry for non-legal professionals, and reduce the costs of license clearance when legal experts are needed to resolve a client request.

2.2. DALICC Software Architecture & Functional Spectrum

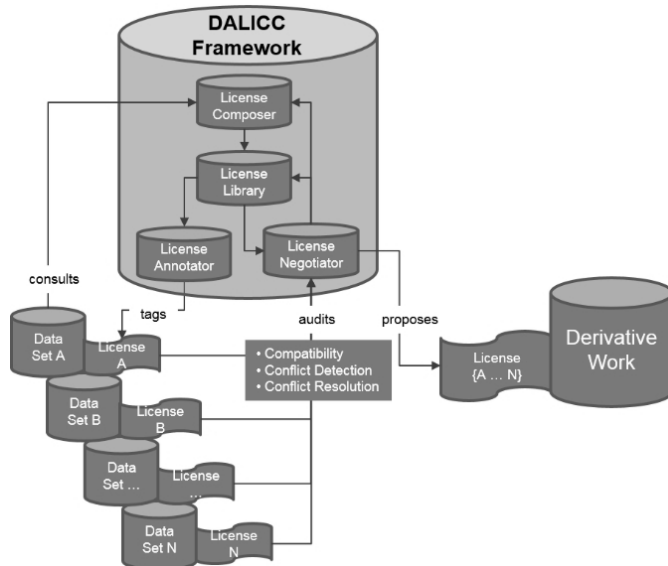


Figure 1: DALICC Software Architecture

To tackle to above mentioned challenges the DALICC framework, which is depicted in Figure 1, will consist of four components:

1) License Composer: A tool that allows licenses to be easily created from a set of vocabularies provided by existing RELs. ODRL will be used as a baseline that will be extended with other REL vocabularies, in order to satisfy the requirements of real world use cases. The composer will not just be a vocabulary editor but also allow to specify ambiguous concepts according to the application provider's usage scenario (i.e. «The non-commercial clause includes direct payments AND advertisement» or «Distribution shall be understood as ...»).

2) License Library: A repository that contains machine-readable and human-readable representations of licenses. Starting from a 80-20-rule, in the first phase of the project the repository will be populated with the most important licenses used in the data domain. According to a recent study by ERMILOV & PELLEGRINI [2015] among these are: CC0, CC-BY, CC-NC, UK-OGL, DL-DE-BY-1.0, IODLv2. In a second phase the repository will be extended with additional licenses that frequently appear in the data domain or are of specific

importance for future applications (i.e. national Open Data licenses). At the end of the project the library shall contain licenses provided by Creative Commons and Open Data Commons, as well as national open data licenses (i.e. Austria, Germany, Italy, UK etc.) and popular vendor specific licenses (i.e. Oracle, Socrata, DataCollect).

3) License Annotator: A tool that allows machine-readable and human-readable licenses to be attached to a dataset or subsets thereof. This can be done by either choosing standard licenses (i.e. CC-BY) already available in the License Library or creating a custom license by using the License Composer. Each newly created license will automatically be added to the License Library thus allowing incremental growth of the repository and the associated knowledge base. The licenses will also be available in various formats.

4) License Negotiator: A tool that checks license compatibility and supports conflict detection between licenses. This will be the core component of the DALICC framework that caters for reasoning over licenses taking into account the specific context of the application provider. The core functionalities will be to detect equivalent licenses being provided under different names (i.e. CC0 ↔ NL Publiek-Domein ↔ PDDL), to point to potential and factual conflicts deriving from non-equivalent licenses and to support conflict resolution. Identified resolution strategies, e.g. for re-establishing compatibility among a set of licenses, shall not solely refer to choosing the most restrictive license at hand hence potentially reduce the usefulness of the resulting (combined) license, but proposing a semantically equivalent and legally sound alternative license representation that resolves detected conflicts.

3. Conclusion

DALICC will contribute its use cases and requirements to the W3C ODRL working group on Permissions & Obligations Expression (<https://www.w3.org/2016/poe/charter>), thus making sure that certain concepts/characteristics make it into the final REL and are immediately adopted and incorporated into the DALICC framework.

The DALICC framework will be available under a dual license, thus allowing various forms of exploitation. The framework closes the existing gap between the technological capabilities to create and publish data, and the legal infrastructure necessary to provide them on a legally secure basis for reuse. Hence, DALICC is a tool that puts data policies into practice and thus facilitates data governance. According to the data value chain provided by DELOITTE [2012], the DALICC framework should be understood as an enabling service for the emerging data economy.

4. Acknowledgements

The DALICC project is a cooperative project funded by the Austrian Research Promotion Agency and takes place in the timeframe of 1 November 2016 to 30 October 2018. Details are available at <http://www.dalicc.net/>.

5. References

- DELOITTE (2012). Open growth. Stimulating demand for open data in the UK. See also <http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf>.
- ENISA (2013). Detect, SHARE, Protect Solutions for Improving Threat Data Exchange among CERTs, October 2013.
- ERMILOV, IVAN, & PELLEGRINI, TASSILO (2015). Data licensing on the cloud: empirical insights and implications for linked data. ACM Press, p. 153–156.
- EUROPEAN COMMISSION (2014). Towards a thriving data-driven economy. Brussels, 2 July 2014, COM(2014) 442 final.
- HOFFMANN, AXEL, SCHULZ, THOMAS, ZIRFAS, JULIA, HOFFMANN, HOLGER, ROSSNAGEL, ALEXANDER, & LEIMEISTER, JAN MARCO (2015). Legal Compatibility as a Characteristic of Sociotechnical Systems. *Business & Information Systems Engineering*, 57(2), p. 103–113. <http://doi.org/10.1007/s12599-015-0373-5>.
- 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.