COMPUTER LAW & SECURITY REVIEW ■■ (2016) ■■–■■

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Legal aspects of linked data – The European framework

Víctor Rodríguez-Doncel [a,*], *Cristiana Santos* [b], *Pompeu Casanovas* [b,c], *Asunción Gómez-Pérez* [a]

[a] *Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain*
[b] *Institute of Law and Technology (IDT-UAB), Universitat Autònoma de Barcelona, Barcelona, Spain*
[c] *Faculty of Business and Law, Data to Decisions Cooperative Research Centre, Deakin University, Geelong, Australia*

Keywords:
Linked data
Open data
Copyright
Database right
Data protection
Licences

A B S T R A C T

This paper portrays a general overview of the existing European legal framework that applies to the publication and consumption of linked data resources in typical settings. The point of view of both data publishers and data consumers is considered, identifying their rights and obligations, with special attention to those derived from the copyright and data protection laws. The goal of this analysis is to identify the practices that help to make the publication and consumption of linked data resources legally compliant processes. An insight on broader regulations, best practices and common situations is given.

1. Introduction

The World Wide Web was born 25 years ago and it has changed the way humans access information. The key of its success lay largely in the general adoption of common practices and their ability to create a network of linked documents or *hypertext*. These practices were formalised by the World Wide Web Consortium (W3C) as "W3C Recommendations", namely, public specifications discussed by a broad community. For example, the version 5 of HTML is a W3C Recommendation published in 2014.

The W3C Consortium, still led by its founder Tim Berners-Lee, has published a new set of W3C Recommendations in the last few years towards the implementation of a "semantic web" of data. Much like the linkability of documents in the World Wide Web, the new web's most distinct feature is the ability to reference chunks of data eventually published by others. This *web of data* (term opposed to the original *web of documents*) is intended to be accessed not only by humans, but also and mostly by machines. The ultimate practices recommended by the W3C to publish data on the web refer to it as "*Linked Data*". Linked data is configuring a global data space of high quality data, strongly interconnected and with peculiar features that make it different from a database from the legal point of view.

For example, machines understand particularly well the information offered as linked data, because data models are typically specified by computer ontologies which enable automated reasoning. Linked data favours connecting distributed pieces of information, relating apparently disparate datasets. Linked data uses crowdsourced vocabularies, where no authority can impose the use of a specific variant. Linked data can be ephemeral, and it is hard to prove that some data was online at a precise time. In sum, a collection of *linked data* is

* *Corresponding author.* Artificial Intelligence Department, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte, 28660 Madrid, Spain.
E-mail address: vrodriguez@fi.upm.es (V. Rodríguez-Doncel).

http://dx.doi.org/10.1016/j.clsr.2016.07.005

Please cite this article in press as: Víctor Rodríguez-Doncel, Cristiana Santos, Pompeu Casanovas, Asunción Gómez-Pérez, Legal aspects of linked data – The European framework, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.07.005

not merely a database and its unique characteristics require special attention.

Linked data is gaining momentum. Many public institutions are publishing their datasets as *linked data* in open data portals, where anyone can download and reuse the information for commercial or non-commercial purposes, unleashing valuable resources for businesses, citizens, and other public administrations.[1] In Europe, the EU Open Data Portal,[2] the cornerstone of the EU open data strategy, provides access to datasets of different institutions (from Eurostats to national weather agencies) with a high penetration of linked data as the publishing format.[3] The *CELLAR* repository, the central component of the information system set by the European Publication Office, uses *linked data* to provide semantic indexing, advanced search and data retrieval for multilingual resources.[4] *Linked data* has also attracted attention in academic publishing,[5] governance development[6] and in the legal domain[7] (including legislation,[8,9] case law[10] and legal education[11]). Industrial organisations are publishing and consuming *linked data* in the operation of their business (as non-open linked data, also called *linked enterprise data*) and new business models are starting to take off, supporting every step of the lifecycle of *linked data*.[12] Thus, standardisation trends have received also a renewed attention, as within the emerging scenarios, rules and principles are at stake with users' behaviour, which may consume, discuss, follow or ignore them.[13]

So far as we are aware there, is no evidence thus far of case-law nor out-of-court disputes regarding *linked data* resources.[14] However, both in EU and USA, there is a large set of case-based decisions already dealing with the unintended effects of knowledge aggregation and profiling – which are capabilities uplifted by semantic web technologies. Conflicts have arisen in different sectors, including financing marketplace, health, and the areas of Freedom, Security, and Justice (which is quite intense in balancing exceptions to liberty rights and security threats[15]). Also, more conflicts are expected to appear as *linked data* is playing a more important role.

The objective of this article is to describe the applicable legal framework in Europe for *linked data*, identifying the applicable rights and obligations in common scenarios and facilitating the *linked data* publication and consumption processes to be legally compliant. The ascribable discipline of rights and obligations in relation to *linked data* arise from legal instruments, such as copyright law, database law, trade secrecy law or data protection law which are briefly examined in this paper.

Section 2 presents linked data resources and features and its role within the Semantic Web. Section 3 analyses the legal framework of linked data in Europe, describing the applicable law with respect to copyright, data protection and other legal statutes, like competition law or trade secrecy law. Section 4 concludes with a review of the overall legal aspects and presents the forthcoming initiatives on the analysed legal framework which shall affect the linked data universe.

2. Linked data

2.1. *Description of linked data*

The definition of *linked data* is often given as the result of adopting *a set of best practices for publishing and connecting structured data on the Web*.[16] These practices can be summarised in:

(i) URIs[17] are used as identifiers of the resources. A resource can be a city, a person or anything there is data about. A URI can be for example http://en.wikipedia.org/wiki/London.

[1] Archer P. et al. (2014). Study on business models for Linked Open Government Data. Technical Report, BM4LOGD, https://joinup.ec.europa.eu/node/72473 (visited April 2016).

[2] The European Union Data Portal is available at https://open-data.europa.eu/. As of April 2016, it accounted 8546 different datasets. A notable precedent was similar intent was the portal http://PublicData.eu, developed by the Open Knowledge Foundation and still online.

[3] As of April 2016, the portal accounted that 46% of the datasets were already machine readable. http://www.europeandataportal.eu/mqa-service.

[4] Francesconi, E., Küster, M.W., Gratz, P., & Thelen, S. (2015). The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services. In Electronic Government and the Information Systems Perspective, pp. 29–39. Springer Int. Publishing.

[5] Peroni, S. Semantic Web Technologies and Legal Scholarly Publishing, vol. 15, LGT Series, Springer, Dordrecht (2015).

[6] Davies, T. and Edwards, D. (2012). Emerging Implications of Open and Linked Data for Knowledge Sharing in Development, IDS Bulletin 43 (5), 117–127.

[7] Casanovas, P., Palmirani, M., Peroni, S., van Engers, T., Vitali, F. (2016). Special Issue on the Semantic Web for the Legal Domain Guest Editors' Editorial: The Next Step, Semantic Web Journal, 7(2): 213–227.

[8] World Legal Information Institute, http://www.worldlii.org/ (visited April 2016).

[9] Casellas, N., Bruce, T.R., Frug, S.S., Bouwman, S., Dias, D., Lin, J. & Venkataraman, S. (2012). Linked legal data: improving access to regulations. In Proc. of the 13th Annual Int. Conf. on Digital Government Research, pp. 280–281, ACM.

[10] The project EU Cases (http://eucases.eu/, visited in April 2016) offers case law as linked data.

[11] Casanovas, P. (2012). Legal crowdsourcing and relational law: what the Semantic Web can do for legal education. In Journal of Australian Law Teachers Association, Vol. 5 (1 & 2) 159–176.

[12] Auer S. et al. (2012). Managing the Life-Cycle of Linked Data with the LOD2 Stack. International Semantic Web Conference (2), vol. 7650 of Lecture Notes in Computer Science, pp. 1–16. Springer.

[13] Polleres A. (2013) Agreement Technologies and the Semantic Web. In S. Ossowski, ed. Agreement Technologies, vol. 8 of LGT Series, pp. 57–68. Springer, Dordrecht.

[14] Search at Lexis Nexis and West Law legal databases made as of mid 2015.

[15] Boehm, F. (2012) Information Sharing and Data Protection in the Area of Freedom, Security and Justice, Springer, Dordrecht.

[16] Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked data-the story so far. Int. J. on Semantic Web and Information Systems, vol. 5(3), pp. 1–22.

[17] A URIs is a string of characters used to identify a resource on the Internet (like mailto:john@example.com or http://example.com). The specification is defined in "RFC 3986 Uniform Resource Identifier (URI): Generic Syntax".

Please cite this article in press as: Víctor Rodríguez-Doncel, Cristiana Santos, Pompeu Casanovas, Asunción Gómez-Pérez, Legal aspects of linked data – The European framework, Computer Law & Security Review: The International Journal of Technology Law and Practice (2016), doi: 10.1016/j.clsr.2016.07.005