

The European Legal Semantic Web: Completed Building Blocks and Future Work

Marc van Opijnen¹

European Legal Access Conference, Paris, 21-23 November 2012

If constructed properly the European legal semantic web will improve access to legal information, stimulate innovative applications and legal services, and reinforce judicial and legal cooperation within Europe.

In this paper we will discuss why we still we do not have one-click answers on very basic legal questions, what building blocks are already in place and what still has to be done to have the European Legal Semantic Web really functioning.

We will start with some illustrations from legal practice to demonstrate the blessings of the semantic web, and the definition of some terminology (§ 1). Next, we will review the state of play regarding the most important building blocks for identifying legal sources (§ 2). In § 3 we will summarize the most necessary steps that have to be taken in the near future, both at European and national level, to make substantial headway. One of these steps might be the development of a European Legal Doctrine Identifier.

Some concluding remarks are made in § 4.

1 Introduction

1.1 Daily Annoyances

Since there is a reasonable chance you are having a legal background, some of the little annoyances outlined below are probably quite familiar to you.

Wouldn't it be nice to have a search engine delivering you not just those judgments which are relevant to article X of regulation Y, but just those judgments that refer to this legislation as valid on a the specific point in time? You might try to use the date of judgment as a filter, but because this is not a fully reliable indicator, you have to filter manually all documents returned. With the help of the semantic web though this search can be performed within seconds.

Or, you are dealing with a difficult case in which the interpretation of a European regulation plays a major role. Before deciding on whether or not to ask the Court of Justice of the European Union (CoJ EU) for a preliminary ruling, you have to know – according to the case law of the CoJ EU² – how this question was dealt with by colleagues in other Member States. But have you ever tried to find case law from abroad on specific European directives

¹ Sr. adviser legal informatics at the Netherlands Council for the Judiciary.

² Court of Justice EU, CELEX:61981J0283 (Cilfit).

or regulations? It can keep you busy (and frustrated) for days.³ On the semantic web though, you'll have your answer within seconds.

One of the cases you did find on the previous search is from some district court. That court didn't ask for a preliminary ruling, but how can you be sure a preliminary ruling wasn't being asked for in appeal? Judicial websites in most countries don't display information on follow-up proceedings. So, you would have to do free-text searches with not very strictly formatted case numbers in search engines that respond in an unpredictable way to the punctuation marks in these case numbers. On the semantic web you would have, just with one click of the mouse, a list of the appeal judgment with the referral to Luxembourg, the preliminary ruling itself, and the follow-up case by the national judge.

Or, you have been confronted with a legal problem of a rare and problematic nature. You came across an interesting foreign decision, and would like to find all relevant scholarly writings on this case, from whatever country. Soon, you'll find yourself dragged into a muddle of incomparable identifiers and different citation styles, leaving you uncertain about the completeness of the collected materials. On the semantic web, you would have all your scholarly writings and other relevant materials listed in seconds.

And finally, not an annoyance but maybe a dream: wouldn't it be nice, starting work on a new case file, to have a complete listing of all relevant legislation (including pending amendments), jurisprudence, recent scholarly writings, related case files and involved colleagues, without having to push any button? Well, this nut is a little harder to crack than the other examples, but, if we invest in constructing the legal semantic web today, this could be your joy tomorrow.

1.2 The European Legal Semantic web

The internet itself brought many advantages for legal professionals: legislation and case law is available for free in substantive quantities and (some) searches are faster. But the user gets lost easily: websites are poorly interconnected, they all have their own interfaces, identification systems and terminology. At least up until recently public data suppliers (like governmental organisations, judiciaries, universities) didn't interpret their task of supplying information any broader than just making the rough materials available, and private companies are generally so overwhelmed by the rapidly changing information market, threatening to undermine their long-standing monopoly, that they shield their data and refuse to participate in initiatives for creating a common framework.

Probably because of the way the legal information market has been dominated by legal publishers for decades, their information architecture has been taken for granted too long, too easily.

With the amounts of free legal information becoming publicly available, the flaws in (or 'the absence of') this information architecture are becoming more and more apparent. Before any major steps can be taken to solve the more interesting challenges on e.g. legal reasoning, these basic architectural flaws have to be fixed.

There is no single agreed-upon definition of the 'semantic web'. We can explain the concept with this short definition by W3C: "*The Semantic Web is about common formats for integration and combination of data drawn from diverse sources, where the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects.*"⁴

Important elements in this definition are:

³ M. van Opijnen, 'Searching for References to Secondary EU Legislation', in *Fourth International Workshop on Juris-informatics (JURISIN 2010)*, ed. by S Tojo (Tokio, 2010). <ssrn.com/abstract=2046255>

⁴ <www.w3.org/2001/sw>

- Common formats: Open standards, agreed upon by open fora, that can be freely used by everyone.
- Data: digital objects that can be everything, not just documents in the traditional sense.
- Integration and combination, necessary to realize functionalities that wouldn't be possible otherwise. There shouldn't be any technical, organizational or legal boundaries on what type of application, organization or domain is using or creating the data.
- Language for recording how data relate to real world objects. This means that the meaning of data has to be expressed in such a (digital) way that computers and humans understand each other very precisely.

While the first three elements (formats, data and integration) are of a more technical nature, the fourth element (language) is of a highly domain-specific nature. For the legal semantic web, this language pertains to how computers and humans can communicate about legal objects and constructs without any misunderstandings. Although in the end we might like to have a legal debate with the computer, the univocal identification of the most essential legal objects would already be a major step forward. If the identification of, and referral to legislation, case law, legal scholarly writings and preparatory and parliamentary documents are dealt with properly, all the one-click solutions conjured up in the first paragraph are less illusive than they might seem at first sight.

While the legal semantic web is that specific part of the semantic web that is about legal data, the European legal semantic web refers to the interconnected and interoperable legal semantic webs of all Member States and that of the European Union. The semantic web is a global phenomenon: it is based on universal standards and has no borders. Domain related semantic webs can be more or less restricted by boundaries of any kind; for the legal domain these limits are quite obvious, because they coincide with jurisdictional boundaries. But with European law becoming such an integral part of the legal systems of its Member States, and with the intertwining of national legal systems as a result of European integration, boundaries between legal systems are increasingly blurred.

With full respect of the competence of each Member State to develop its own building blocks, the development and use of European standards and services will facilitate the functioning and interoperability of all these national legal semantic webs.

2 Objects in the European Legal Semantic Web

2.1 Introduction

In this paragraph we'll discuss the most important legal objects that have to be identified for the legal semantic web to function. We will give an overview of the challenges to be addressed when properly identifying these objects, as well as the initiatives that were already developed to build the required components.

Before turning to our legal objects though, we have to introduce some more terminology. When talking about (legal) bibliographic objects, misunderstandings are lurking if we do not exactly define on what abstraction level we are communicating. Humans are well able to distinguish between the different concepts of 'ruling XYZ' in "*This type of evidence was permitted by the court in its ruling XYZ*" and "*I had ruling XYZ on my desk yesterday, who took it?*" A computer though cannot by itself grasp the differences between these levels of abstraction. Therefore we have to use a classification system to properly distinguish between these ontological levels.

A useful method, developed in the bibliographic world,⁵ uses four basic levels: work, expression, manifestation and item (see box 1). Also, we define ‘identifier’ as ‘a unique number given to any bibliographic object’, and a ‘citation’ as ‘the representation of an identifier, with the intention of referring to the bibliographic object.’⁶ E.g.: an ISBN printed on a book is an identifier, but when you use it in a text to refer to a book which is identified by this ISBN, it is called a citation. Finally, by the term ‘metadata’ we refer to descriptive data about content (e.g. the date of a judgment, the type of a document, the issuing institution).

2.2 Legal documents of the European Union

From the start of the European Communities most basic legal materials were stored in the CELEX-database.⁷ It started with the internal and external treaties, secondary legislation, preparatory documents, case of law of the Court of Justice and parliamentary questions. CELEX was gradually replaced by EUR-Lex, and more document types were included, like all publications from the Official Journal, consolidated legislation and national implementation measures. All these documents have a CELEX-number. Although seen from modern perspective the CELEX-numbering system has some minor flaws,⁸ it is a sound way for uniquely and persistently identifying EU legal sources: it is compact and meaningful, its syntax is error-proof, language-independent and technology-neutral.

The CELEX-number is derived from other document numbering systems used within the Union, and the use of those other numbering systems is prescribed by the Interinstitutional style guide.⁹ Therefore the CELEX-number never gained much acceptance among lawyers. Unfortunately, because the other numbering systems have formats that are verbose, error-prone and language-dependent. In a paper world formatting errors in a citation are irritating but not crucial: as long as lawyers

Box 1: Bibliographical levels

Work: A distinct intellectual or artistic creation. The ‘work’ is an abstract level, it describes only the creation as such. For a judgment it is the judicial decision resolving the specific legal dispute brought before the court, based on a specific set of arguments. This work level is addressed when one says: “*According to the decision of the European Court of Justice in case C-299/02 (...)*.” For legislation it is e.g. ‘the Criminal Code of Sweden’, without any specification on the date of the referral or the type of edition.

Expression: The intellectual or artistic realization of a work. The expression is also an intellectual or artistic product, but it is always derived from the work. Like the work, the expression is also an abstract level. For a judgment an expression could be the original wording by judge or clerk, or the summarized and annotated version by a specific legal publisher. For a European regulation all language versions are distinct expressions. But also temporal editions are expressions: the Criminal Code of Sweden as in force on 11 November 2011.

Manifestation: The physical embodiment of an expression of a work. Although the manifestation is a physical embodiment, it is only a specific type of embodiment. A case law example could be the PDF-version of the Italian expression of a specific ECJ-judgment.

Item: The single exemplar of a manifestation. This could be a PDF case law document residing in a specific directory on my computer.

⁵ International Federation of Library Associations and Institutions, ‘Functional Requirements for Bibliographic Records’, in *UBCIM Publications – New Series Vol 19*, (1998).

<www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf>

⁶CEN, ‘CEN Workshop Agreement Metalex (Open XML Interchange Format for Legal and Legislative Resources)’, <www.cen.eu/cen/Sectors/Sectors/ISSS/Activity/Documents/CWA15710.pdf>.

⁷ On the full history of CELEX: *25 Years of European Law Online*, (OPOCE, 2006), <bookshop.europa.eu/en/25-years-of-european-law-online-pbOA7606136>.

⁸ One of them being the deviant numbering of the internal treaties in sector 1. Also it doesn’t have a fully documented and flawless system for identifying particles of secondary legislation. See footnote 3

⁹ Europese Unie, *Interinstitutionele styleguide*, (Luxemburg: Publications Office of the European Union, 2011). <publications.europa.eu/code/en/en-000100.htm>

understand what is actually meant, citation can be followed. But computers are far less keen on guessing what is meant, and search engines are particularly unfriendly towards punctuation marks used within citations like ‘Regulation (EEC) No. 1408/71’.¹⁰

And although at first sight the CELEX-number might look complicated, it is easier to apply properly than are all those different citation styles for directives, regulations, framework decisions, a.s.o.. And because all information is already in the number itself, the notation saves a lot of space compared to document numbers: ‘CELEX:31971R1408’ suffices. Although better alternatives are now available for legislation (below, § 2.3) and case law (below, § 2.4) the CELEX-number is still by far the best identifier for e.g. Commission documents and documents published in Official Journal.

2.3 National Legislation

Nowadays national legislation is available in freely and publicly accessible databases in most Member States. While it often started with a mere publication of legal gazettes, databases with current and historic consolidated versions are increasingly being developed. The inline hyperlinking between various (paragraphs of) law, but also the need to allow for incoming external references, forced Member States to develop identification systems. Although differences exist in the way identifiers have been built and metadata on legislation classified, they are all based on the same legal concepts.

Cross-border access to national legislation within the EU is becoming increasingly important. Judges and legislators have to take note of the way European directives are implemented in other Member States, and comparative law studies require topical access to foreign legislative databases. To meet this information need the N-Lex portal¹¹ was launched in 2006. It offers a basic tool for accessing foreign legislative databases. Also, EUR-Lex in ‘sector 7’ now gives access to national implementation measures. These initiatives though reveal the problem of all Member States having differences in document structures, metadata and identification systems.

Inspired by the work on ECLI (infra, § 2.4) the EU Council working party on e-law developed the European Legislation Identifier (ELI), adopted by the Council of Ministers on 24 September 2012.¹² Put simply, ELI is a framework which allows for formatting national legislation in a univocal European format. Because it also contains a scheme for the most essential metadata, it allows for a substantial improvement of cross-border access and reuse of legislative data. Although ELI is a single European framework, it allows for national specificities. Therefore every Member State participating in ELI (on a voluntary basis) has to appoint a ‘national ELI-coordinator’, responsible for establishing the specific URI- and metadata schemes. For the EU the Publication Office is responsible for the implementation of ELI in EUR-Lex. The Publications Office will also maintain the register of the Member States’ implementation schemes.

ELI is an identification system for different ontological levels. It identifies the work, and (optionally) a temporal or linguistic expression. Also, it can make a distinction between original act and consolidated version, and it can identify a full law, or just a (very) specific part of it. Finally, it’s important to notice that ELIs are based on “‘*http URIs*’ to specifically identify all online legal information officially published across Europe”. In other words: an ELI always refers to an official publication; it has one single point of truth. As we will see, in (i.a.) this respect it differs from the identification of case law.

¹⁰ See footnote 3.

¹¹ <n-lex.europa.eu>

¹² ‘Council conclusions inviting the introduction of the European Legislation Identifier (ELI)’, CELEX:52012XG1026(01).

2.4 Case Law

Case law as an object in the legal semantic web differs from legislation in two important respects.

First, while legislation has always been meant to be public, case law is meant to settle disputes between parties, and is therefore – in general – not meant to be published in full. Case law is only published when it has legal or societal relevance outside the boundaries of the dispute decided. As a result, various expressions of one judgment exist: anonymized, shortened, summarized or translated. Unlike legislation, where only the official text is the ‘true version’, for case law many of these ‘editorial expressions’ exist. Although the version on a ministerial or judiciary website can be considered being ‘most authentic’, this might not be the one preferred by users, and it might not even be available: some judgments are only published elsewhere or removed from the official website after a fixed period.¹³ The existence of these editorial expressions requires a substantial different approach for the construction of a work level identifier and the way expressions are linked to the work.

This touches upon the second difference between legislation and case law: although in many countries private legal publishers had their share in the publication of legislation, they have been true monopolists in the publication of case law. Specific periodicals published selections of interesting cases, which were all identified by a publication number. In our ontological classification these are ‘expression identifiers’; one decision could easily have ten different expression identifiers, without revealing that they actually related to the same work. The only work identifier was the combination of court name, case number and decision date (hereafter referred to as ‘triple’). In some Member States specific judgment identifiers were developed for cases published on governmental or judicial websites,¹⁴ but they are not always linked to commercial expression identifiers and are not used very often by lawyers for citation purposes. Judges in their decisions, and academics in their scholarly writings keep using triples and expression identifiers to cite jurisprudence. As a result, hyperlinks cannot be constructed, citation indexes cannot be built automatically,¹⁵ case law cannot be found, legal knowledge on e.g. questions of European law is not shared in a way that suits the requirements of the CoJ EU and other European Institutions,¹⁶ and legal and judicial cooperation within the European Union is seriously hampered.

These differences between the nature of legislation and case law, are reflected in the differences between ELI and the ECLI, the European Case Law Identifier. The EU Council Conclusions on the European Case Law Identifier¹⁷ were adopted in December 2010. The technical annex of these conclusions include specifications on the way in which this identifier has to be constructed.

Unlike the ELI, which can have up to fifteen (or more)¹⁸ constituting elements, ECLI always has five,¹⁹ separated by a colon:

¹³ E.g. at the websites of the German ‘*Bundessozialgericht*’ <www.bundessozialgericht.de> and ‘*Bundesarbeitsgericht*’ <www.bundesarbeitsgericht.de> only the decisions of the running and four preceding years can be accessed.

¹⁴ E.g. ‘*Repertorio Oficial de Jurisprudencia*’ (ROJ) in Spain and ‘*Landelijk JurisprudentieNummer*’ (LJN) in the Netherlands.

¹⁵ With a case law citation index a user can get an instant overview of all – preferably hyperlinked – judgments and scholarly writings in which a specific judgment has been cited.

¹⁶ E.g. European Parliament, ‘European Parliament resolution of 9 July 2008 on the role of the national judge in the European judicial system, CELEX:52008IP0352.

¹⁷ Council of the European Union, ‘Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law’, CELEX:52011XG0429(01).

¹⁸ Dependent on the number of sublevels that is being identified.

¹⁹ For a more elaborate discussion on ECLI see e.g.: M. van Opijnen, ‘European Case Law Identifier: indispensable asset for legal information retrieval’, in *From Information to Knowledge. Online Access to Legal*

- ‘ECLI’ as a self-identifier;
- a two-letter country code;
- a code for the name of the court;
- the year of judgment (not being the year of registration, which is often part of the case or docket number);
- a unique identifier, to be decided by the implementing country or organisation.

Implementation of ECLI already started i.a. in France and Slovenia, both using case number and date information in the fifth part, resulting in the respective examples:

ECLI:FR:CESJS:2012:330962.20120822 and ECLI:SI:VSLJ:2011:II.IP.1301.2011. The CoJ EU will use the CELEX-number for the fifth part, e.g. ECLI:EU:C:2009:62005CJ0065.²⁰

Of course, one has to get used to the format of this identifier, but the advantages are clear. To name a few:

- One never has to question the nature of the identifier. The fact that it starts with ECLI makes it recognizable by lawyers across Europe immediately;
- The next three parts reveal a lot of extra information. One knows immediately in which country the decision is taken, in which year and at which court. The court codes of one’s own country will be remembered soon enough, the court codes of other countries will be available on the internet.
- ECLI might seem a long identifier, but since it contains all relevant information one doesn’t have to add dates, name of court or expression identifiers.

The last point is crucial and needs some explanation. While ECLI refers to one web address, leading to the legislative text itself, an ECLI in itself doesn’t give any clue on where to actually find the text of the judgment. This function will be implemented by the ECLI-search interface, also decided upon in the Council Conclusions. This ECLI search interface will be part of the European e-Justice Portal,²¹ and will enable both public and commercial case law databases to register any expression of a judgment having an ECLI. Via this search interface the end-user can have an immediate (hyperlinked) overview of all locations where a judgment is published. And since the Council Conclusions also prescribe a set of (optionally quite comprehensive) metadata, judgments will also be searchable by summaries, translations, keywords a.s.o. The ECLI search interface will also offer the opportunity to search the documents themselves (if available).

The European Commission is responsible for the implementation of the ECLI search interface. The implementation of ECLI on the national level though is a national responsibility. A national ECLI-coordinator is responsible for assigning court codes and documenting the way the fifth part is constructed.

For best performance, ECLI should be implemented at all courts in all Member States, for all judgments (not only those published), both historic and future. To avoid any barriers for participation though, the ECLI system is voluntary, and implementing Member States are free to assign ECLI only to new cases, to start with just some courts or with the whole judiciary at once, to use ECLI as an additional number next to other identifiers, or to use ECLI also as the primary national identifier, to use ECLI only on published cases or for all case law.

Information: Methodologies, Trends and Perspectives, ed. by M A Biasiotti and S Faro (IOS Press, 2011), p. 91-104. <ssrn.com/abstract=2046160>

²⁰ B. Gielen and N. Lieber, ‘Implementation of the ECLI project at the level of the European Court of Justice’, <www.aca-europe.eu/seminars/Warsaw2011/Exp_Gielen_Lieber.pdf>.

²¹ <e-justice.europa.eu>

2.5 Legal Terminology

So far we discussed legal sources, but a proper functioning of the legal semantic web requires also a computer readable language for defining legal terms and concepts. This is a far more complex and challenging job than the identification of legal sources. Although having different numbering systems, different suppliers and different metadata schemes, legal sources like legislation and case law are well-defined, identifiable and limited. Lawyers will agree on what is meant by ‘decision X’ or ‘paragraph Y of regulation Z’, even on the various ontological levels. But ask them to draw up a list of the most important legal terms, or ask them to define ‘tort’, and they will discuss infinitely. But if we want the legal semantic web to be able to supply all documents on tort, the computer has to know which documents to retrieve – while just retrieving all documents containing this term will not suffice. And although sometimes a basic mutual understanding of terms exists,²² more refined data integration functionality requires a more precisely described legal vocabulary.

Some building blocks already exist. Eurovoc²³ is a multilingual thesaurus, covering the activities of EU. It is now also available in SKOS/RDF-format, which makes it usable in semantic applications. The InterActive Terminology for Europe (IATE)²⁴ is the EU inter-institutional terminology database, containing approximately 1.4 million multilingual entries. For legal purposes it is of better use than Eurovoc, but is not as ready for the semantic web.

A high-level semantic framework for the exchange of legal information was developed with the ‘Legal Knowledge Interchange Format’ (LKIF),²⁵ a deliverable from the Estrella-project.

More work is done in a great variety of projects, e.g. in e-Codex (E-justice Communication via Online Data EXchange),²⁶ aimed at developing building blocks that can be used in and between Member States to support cross-border proceedings in the field of justice.

3 Future work

As discussed above, important work has been done in recent years to bring a European Legal Semantic Web closer to reality, for the benefit of European legal professionals, governments, judiciaries and citizens. But although the steps in themselves have laid an important foundation, we still cannot supply the average user with the one-click search actions listed in the first paragraph. Therefore, further steps are needed. Some of the most fundamental ones are outlined below.

3.1 Integral Approach

While in the Multi-annual European E-Justice Action Plan 2009-2013²⁷ access to information in the field of justice²⁸ is defined as one of the three basic functions for the future European e-Justice system, the Action Plan itself mainly focuses on the other two functions: dematerialisation of proceedings and communication between judicial authorities. With the Action Plan reaching the end of its five-year term and a continuation being discussed, more attention should be drawn to the European legal semantic web. Although judicial cooperation and on-line proceedings also have a lot of semantic issues involved, the European legal

²² E.g. in the ECLI-metadata basic fields of law can be assigned to a judgment

²³ <eurovoc.europa.eu>

²⁴ <iate.europa.eu>

²⁵ <www.estrellaproject.org/?page_id=5>

²⁶ <www.e-codex.eu>

²⁷ Multi-annual European E-Justice Action Plan 2009-2013', CELEX:52009XG0331(01).

²⁸ “This information concerns in particular European legislation and case law as well as that of the Member States.” (§ 26 of the Action Plan).

semantic web has its own specific requirements, and it requires a more integral approach. Before discussing some more specific elements, some organisational issues have to be raised.

First of all, it should be made more explicit why European action is needed. Recent history has shown that most Member States struggle with the architecture of a legal semantic web at the national level. This is not only due to the complexity of the data, but also the variety of players involved: legislator, judiciary, universities and commercial legal publishers all have their own interests to pursue, complicating and hampering the development of open standards. The establishment of (even voluntary) standards at the European level might end deadlocks and stimulate national developments. A second reason for European action is the already discussed growing interdependency between European and national legal sources and systems.

Second, and this might not be stressed enough, Member States must be able to profit from a shared semantic architecture, while keeping the liberty to organize things at the national level according to their own needs. Both ECLI and ELI leave it to the Member States to use these standards parallel to national standards, to encapsulate national standards or to completely replace national standards. Such a decentralized approach, profiting from the best of both worlds, is also laid down in the current European E-Justice Action Plan and is a cornerstone of the European Interoperability Framework.²⁹

A third element requiring attention is the division of the mutual responsibilities within the EU. The variety of tasks is too broad for a single institution to be fully responsible, but good cooperation is essential to achieve common goals and optimal functionality for all contributors and users.³⁰

3.2 Completion of Identification Standards: the Need for a European Legal Doctrine Identifier

With CELEX-numbers, ECLI and ELI established, some of the most basic legal objects in the legal semantic objects can be identified uniquely and persistently. Apart from the already mentioned legal terminology, requiring continuous attention, one important resource is still unmentioned: legal doctrine.

Although the boundaries between the legal fields and adjacent scientific fields are sometimes blurred, legal doctrine is, in general, quite distinct from those other domains. Most legal doctrine is published by specialized publishers, in specific legal (paper or electronic) magazines or websites. Although not as advanced as in other domains, open access is also becoming increasingly popular within the legal field. For the identification of documents – or more granular information – this is a complicating factor. One article can e.g. be published in a printed magazine, be available for subscribers on the website of the magazine, be published in the open access repository of the academic institution, be published on the writers' personal website, and be uploaded on SSRN.

Although one can refer to each individual manifestation,³¹ using even identifiers as DOI³² or using a standard like OpenURL to address the article, there is no single, univocal way of identifying legal doctrine at the work level. Constructing such an identifier is complicated by the broad variety of forms in which legal doctrine is organized and published. On the one hand there are one-off publications, like PhD-theses, law review articles, case law annotations and conference papers. In principle, such publications are clearly identifiable

²⁹ <ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf>

³⁰ E.g., confusion might arise from the different websites where information has to be gathered. Information on national ECLI-implementation has to be collected at the e-Justice portal, while information on ELI is to be found on EUR-Lex.

³¹ 'Manifestation' as defined in box 1.

³² Digital Object Identifier, <www.doi.org>

objects. But on the other hand one finds ‘permanent’ publications, which are quite specific for the legal domain. These are continuously updated works on the status, interpretation and implementation of the law, aiming constantly to reflect the actual status of affairs in a specific field of law. Previously they were vast paper publications, nowadays they are mostly in an electronic format. Because the content of such websites is changing constantly, legal accuracy requires one to be able to cite a specific temporal expression of the work, at specific granular levels, e.g. a whole chapter or just a specific subparagraph.

A European legal doctrine identifier (ELDI) could possibly help solving this still very problematic issue in the construction of the legal semantic web. An ELDI-architecture should contain elements both of ELI (temporal expressions, granularity), as well as ECLI (a non-http-bound URI and a public register). Of course an ELDI would have to be written in a user-friendly and computer-readable format. Basic metadata, like the locations where to find the information, the year of publication, the type of document, original publisher, author information and optional data like a summary or keywords should be part of an ELDI-standard.

3.3 Implementation of Identification Systems

Of course, just having identification standards is not enough for the European legal semantic web to function. Standards have to be implemented, and often this is more problematic than establishing the standards themselves. At the national level decisions have to be taken on the attribution of tasks, technical schemes developed, databases and applications adapted or rebuilt, and changes communicated to legal professionals and other stakeholders. The work to be done varies per Member State, but a full implementation often requires a process of years. Continuous monitoring, stimulation and the exchange of best practices is therefore needed, and a European action plan could play an important role in this process.

3.4 Implementation in Citation

The correct citation of legal sources can be a lawyers’ nightmare. In the United States law students are being prepared for a life-long struggle with the (in)³³ famous Bluebook,³⁴ in many European jurisdictions citation guides also exist,³⁵ but compliance is less mandatory. Specific guides on citing foreign legal sources also exist,³⁶ but they are generally unknown, or at least not complied with. The European Union has its own style guide,³⁷ in the complexity of which a user might easily drown; many national and language-specific habits and oddities have wrestled their way into this guide in such a way that semantic interoperability can be seriously hampered.³⁸

Apart from these citation guides also legal training cultivates citation habits that are not compliant with the semantic web. And for all information to be integrated fully in the semantic web it shouldn’t just have a proper identifier itself, but also cite all other sources using the right identifiers.

³³ According to i.a. R.A. Posner, ‘The Bluebook Blues’, in: *Yale Law Journal*, 120 (2011): “*The Bluebook (...)* exemplifies hypertrophy in the anthropological sense. It is a monstrous growth, remote from the functional need for legal citation forms, that serves obscure needs of the legal culture and its student subculture.”

³⁴ *The Bluebook: A Uniform System of Citation*, 19 edn (Harvard Law Review Association, 2010),

³⁵ E.g. in France – S. Cottin, *La gestion de la documentation juridique*, (Lextenso éditions, 2011) – or in the Netherlands – G.A.I. Schuijt and others, *Leidraad voor juridische auteurs 2010* (Deventer: Kluwer, 2010).

³⁶ *Guide to foreign and international legal citations*, 2006 (New York University, 2006).

³⁷ See footnote 9.

³⁸ Just to give an example: articles inserted by amendment are having ‘a’, ‘b’ and ‘c’ as their suffix in i.a. English and German language versions of regulations and directives, ‘bis’, ‘ter’ and ‘quater’ in i.a. French, Italian and Dutch language versions.

To solve this very important issue two lines of action are possible. Both require investments, but of a completely different nature.

The first policy line is to change the way in which legal objects are cited by lawyers themselves. Citation guides have to adapt (or adopt) as to give preference to CELEX, ECLI and ELI above any other (expression) identifier, universities have to teach their law students new ways of citation, and already practising lawyers have to be convinced that the new ways of citations are necessary. Trying to achieve these goals on a voluntary basis might be quite cumbersome, and therefore two supportive measures can be imagined.

First, the European identifiers could be made the main identifiers at the national level, discouraging the use of other identifiers. For ECLI this policy will e.g. be followed in the Netherlands, where all case law published by the judiciary will have an ECLI as identifier, and legal publishers will have the opportunity to receive an ECLI for those decisions that haven't been published by the judiciary itself.³⁹

A second supportive measure could be a legal instrument, obliging judges in their decisions, and lawyers in their writings to the court, to use the correct identifiers. For such a measure to be effective, the identifiers must have been fully implemented and drafting software adapted.

A second line of action, which can be combined with the first one, is technical in nature. Here, lawyers can continue citing in their clumsy, erroneous ways, but software is developed that recognizes all these citations and converts them to CELEX-numbers, ELIs and ECLIs. For various types of citations such pattern recognition and canonicalization software has already been developed.⁴⁰ Because of differences between national citation styles such software cannot be easily reused in other jurisdictions, but this type of software isn't rocket-science either.

Even in case the first line of action is chosen, and lawyers are persuaded (or forced) to change their citation habits, such conversion software might be necessary to disclose historic collections.

3.5 Availability of Data

For the legal services and legal information market to function optimally data have to be available as wide as possible. This enables reuse and stimulates the development of new applications. In this respect, the open data initiatives of many Member States and the EU are extremely relevant. It is noteworthy though that in most policy documents on 'Linked Data', legal data are hardly a priority domain, while both the amount, market value and societal relevance of these data should give cause to the opposite.⁴¹

Data sets of legal data should be made better available: free of charge, well-identified, and with all mark-up and metadata that are available.

³⁹ ECLI will be implemented in the Netherlands in 2013. There is already a work level identifier for case law, instituted by the judiciary, assigned to (nearly) case law published in the Netherlands, both in public and private databases. A register containing the work level identifiers, triples and expression identifiers is available for the public (see: M. van Opijnen, 'A Public Index of Case Law References – the End of Multiple and Complex Citations', in *Legal Knowledge and Information Systems - JURIX 2006: The Nineteenth Annual Conference*, ed. by T M van Engers (Brussels: IOS Press, 2006), p. 71-80. <ssrn.com/abstract=2046297>).

⁴⁰ A parser that detects triples, expression level identifiers and work level identifiers, and converts them all to one canonicalized format is described in: M. van Opijnen, 'Canonicalizing Complex Case Law Citations', in *Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference*, ed. by R Winkels (Liverpool: IOS Press, 2010), p. 97-106. <ssrn.com/abstract=2046274> Detection and conversion of references to European secondary legislation is described in the paper mentioned in footnote 9.

⁴¹ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Open data, an engine for innovation, growth and transparent governance, CELEX:52011DC0882.

4 In Conclusion

Important steps have been taken already in building the European Legal Semantic Web. Identification and metadata standards are developed for the most important legal sources. CELEX-numbers, ELI and ECLI could be supplemented with an identifier for legal doctrine, and also on legal terminology work is still to be done.

In the long run even a more mandatory character of the current (and possibly) future instruments might have to be considered: a point could be reached where non-compliance to the standards of the European legal semantic web is considered to be an obstacle for the proper functioning of the internal market for legal (information) services.

But with establishing standards though the work just started. Identifiers have to be assigned to millions of documents, business processes changed, software adapted and developed, citation habits of thousands of not very tech-savvy lawyers changed. And those responsible for accomplishing these actions have to be convinced that the financial investments will pay out in the not even too long run. With a sensible policy and cross-border co-operation these investments might not even be that problematic.

But for all this European orchestration is inevitable. Otherwise the European Legal Semantic Web will end like a house without a roof: built on a solid foundation, but not fit to live in.