Opening Public Data: a path towards innovative legal services

Tommaso Agnoloni, Maria Teresa Sagri, Daniela Tiscornia [*]

[*]*Institute of Legal Information Theory and Techniques of the Italian National Research Council (ITTIG -CNR)*
{agnoloni,sagri,tiscornia}@ittig.cnr.it

Abstract The Open Data movement has gained momentum during last year under the impulse of initiatives on government transparency carried on in different countries starting from the U.S. and the U.K. with the publication of the public data portals data.gov and data.gov.uk. Legal information has not much been interested by this phenomenon so far. We argue that the adoption of the linked data principles for publication of legal data, joined with existing efforts of standardization in the identification and representation of legal information, would open the way to a whole range of innovative legal services and applications based on top of a "Legal Data Cloud". A case study on relevant European legal datasets is presented.

Keywords: Linked Open Data, Open Government, Legal Data Cloud, Public Sector Information, Legal services

1. Introduction

In a technical sense, the term *open data* refers to the 'open' format with which digital data can be distributed on the web to make them more accessible, reusable and interconnected. Broadly speaking, *Open Data* constitute one of the most relevant innovations in the Internet world, because, by enabling the immense non-structured knowledge of the web to be processed and 'consumed', an unlimited panorama of applications is opened up. On data published in open formats it is possible to create services, to build by means of *Linked Data,* new knowledge repositories, and to enhance the path towards a full free access to digital knowledge.
The Open Data paradigm, which has taken on the characteristics of a real movement, has a direct impact on a further relevant reality of Internet, that concerning the *re-usability of Public Sector Information (PSI).* Under the stimulus of the European Union Directive (2003/98/EC), the trend throughout Europe is rapidly growing for both local and national public administrations to publish their data. The scope of institutional producers is two-fold: on the one hand, to offer re-users the possibility of exploiting the maximum benefits from PSI in term of economic opportunities; on the other, to fully realise the paradigm of *Open Government,* that is to say, the accessibility and transparency of public administrations that acquires a completely new meaning in this context. A further important point is the economic benefit

inside the administrations themselves in terms of costs recovery, improved efficiency in data handling and free services providing. The aim of the contribution is to highlight the potentialities of the phenomena in enhancing free access to legal data and legal information services.

2. The Open data initiative

The open data phenomenon is linked to the evolution of web technologies and the ever increasing potentialities offered by interlinking data in order to create new knowledge and new services. It is a relatively recent phenomenon, but one with great impact on the information market. It arises out of an innovative approach to data management originating in the academic world (Semantic Web languages) and it translates into a revolutionary and immediately operational vision of the Internet.
For *"Open Data"* we generically mean data expressed in a platform independent format, machine readable and made available to the public without preventing *reuse* of the conveyed informative content.
From a technical point of view, these requirements are commonly guaranteed, for the representation of data, by the adoption of XML technologies whose specifications are defined by the *W3C Consortium,* a independent non-profit body responsible for the definition of all the technical specifications on which the World Wide Web is based.
The availability of data in XML format is a fundamental step in guaranteeing the syntactic interoperability of data coming from heterogeneous sources and, consequently, in ensuring applications interoperability. This allows data to be freed from the use of specific software products or operational platforms and allow the content of data (its informative content) to be separated from its presentation (aspects that are only hardly decoupled in case of publication of pdf documents or html pages). Access to the *"raw data"* is, in fact, a fundamental requirement for freely reprocessing and reusing it in contexts different from those where the data was originally published.
In order to encourage as much people and institutions to make data available in open format, a sort of rating system of increasing commitment and quality has been set up (Berners-Lee, 2006). Data should be:

1. available on the web (whatever format), but with an open licence useful for their reuse by third parties
2. available as machine-readable structured data (e.g. txt, html or xml instead of pdf or image scan of a document)
3. as above plus non-proprietary format (e.g. odt instead of doc)
4. all the above plus: use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at data
5. all the above, plus: link data to other datasets to provide extended context.

Along side the phenomenon of Open Data, there has been a rapid explosion of the *"Linked Open Data" (LOD)* movement owing to the gradual success

of *Semantic Web* technologies, in particular through the spread of the RDF (Resource Description Framework) model, that constitutes the initial step towards the semantic web.
Entirely based on current web technologies (HTTP, URI, HTML) in accordance with a "layered" standardisation model where every level of the layer is complementary and compatible with the level underneath in order to guarantee a gradual adoption of the specifications, the RDF model provides the link between data from heterogeneous sources by introducing the missing relational level for interconnecting data, namely, *Linked Open Data.* In the words of Tim Berners Lee, inventor of the Semantic Web and one of the principal driving forces behind the LOD movement, the evolution of Internet is represented by a transition from a 'web of documents' towards a 'web of data'.
Just as today's web is made up of the global interconnection of "documents" thanks to the spread of html pages and of hyperlinks that allow for their connection, the vision of the web of the future is, in fact, that of a global interconnection of "data", or (digital) resources in general, unequivocally identified through URI (Unique Resource Identifier) and linked by assertions expressed in RDF. The basic idea of Linked Data is in fact to apply the same architectural principles of the classic document Web to the task of sharing structured data on global scale by combining simplicity with decentralization and openness.

From a technical point of view *Linked Data* refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were first introduced by Tim Berners-Lee in his Web architecture note *Linked Data* (Berners-Lee, 2006) and have become known as the *Linked Data principles*.

According to such principles basic requirements for the publication of linked data on the web are

- The use of URIs for identification
- To expose the data for access via the HTTP protocol
- The use of the RDF data model to describe content of resources and to link them to other useful information

The RDF model encodes the relations among data in the form of assertions made up of "triples" of the kind (*subject, predicate, object)*. The subject and the object of an RDF assertion are both URIs that identify any digital resource. The predicate specifies *how* the subject and object of a triple are related. In this way, the RDF triple may be seen as a (*typed*) link between data coming from different data sources in the same way as links between hypertexts (*untyped*) connect an HTML page to another on the Web. The RDF language, for example, allows assertions on its relevant properties (metadata) to be associated to a data in a format that can be processed by a program. For example, an RDF triple can assert that two resources, a *person* A and a *document* B, both identified by a URI are linked by the fact that A *is*

the author of B. These resources can be exposed on the web in distributed data bases creating in this way a *Web of Data*.

The key advantage of the linked open data model is therefore interoperability: the value of data significantly raise when different datasets created and published independently by different stakeholders can be freely reused and correlated by third parties without technical barriers. This is the base for the creation of added value over data: vertical applications accessing this global distributed dataspacace to offer inedited views and services.

Owing to its simplicity and thanks to its full compatibility with pre-existing informative layers, the adoption of the Linked Open Data paradigm opens new scenarios in all fields related to knowledge and, in particular, in the field of public sector given the vastness and intrinsic value of data gathered and stored by public institutions. In this sense, for *Government Data, Public Data* or *Public Sector Information (PSI),* we mean all the data produced, gathered or stored by public institutions.

3. Open Government Data

The application of the technological paradigm of Open Data and Linked Open Data to public data is an important enabling factor for the full realization of the *Open Government* model. Open Government[1] is, above all, a doctrine according to which an administration has to be transparent on all levels and should permit continual checks on its operation through the use of new technologies. The motivations behind *Open Government Data* initiatives (aimed at making public data freely usable, re-usable redistributable to any one) are, therefore:

– *Transparency*. In a democratic society, citizens need to know what their government is doing so they can exercise public control on its operation. To do that, they must be able to *freely* access to information and public sector data *and* to share that information with other citizens. Transparency isn't just about access but also sharing and re-use. Often to understand data, it is necessary to analyse them, putting them into relation with other data and visualize them in different ways and this requires that they are *open* and can be freely *used* and *re-used.*
– To generate *economic and social value.* In the information society, data represent a basic asset and their value is obviously also an economic one. To make public sector data public means creating, on the one hand, a new business opportunity based on data re-use, and on the other, innovative services for citizens.
– To encourage new forms of *democratic participation*. The Open Government paradigm based on the free and open availability of raw data reprocessable by users is largely influenced by the Opensource software

[1] See, on this http://opengovernmentdata.org/ a project of theWorking Group on Open Government Data at the Open Knowledge Foundation (http://okfn.org/).

movement principles, where the source code of the programs is freely accessible and re-usable by developers to create new applications. Open Government introduces new possibilities of interaction between citizens and institutions and allows new forms of citizen involvement in public life.

In 2007, a working group of supporters of *Open Government in the United States* proposed a list of principles which has become the *de facto* standard for evaluating the opening of government data. The publication of data by public bodies must conform to the following principles:

1. *Data Must Be Complete* - All public data are made available. Public data are data that are not subject to valid privacy, security or privilege limitations.
2. *Data Must Be Primary* - Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.
3. *Data Must Be Timely* - Data are made available as quickly as necessary to preserve the value of the data.
4. *Data Must Be Accessible* - Data are available to the widest range of users for the widest range of purposes.
5. *Data Must Be Machine processable* - Data are reasonably structured to allow automated processing.
6. *Access Must Be Non-Discriminatory* - Data are available to anyone, with no requirement of registration.
7. *Data Formats Must Be Non-Proprietary* - Data are available in a format over which no entity has exclusive control.
8. *Data Must Be License-free* - Data are not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.
9. *Compliance must be reviewable.* An administrative or judicial court must have the jurisdiction to review whether the agency has applied these principles appropriately.

4. Linked Open data and Legal Information

Among public sector information, legal information availability is of primary relevance to enable informed citizenry. From the Declaration on Free Access to Law: "*Public legal information from all countries and international institutions is part of the common heritage of humanity. Maximising access to this information promotes justice and the rule of law;*"[2].

Indeed, nowadays the amount of available unstructured (or poorly structured) legal information and documents made available as part of public accessibility projects by governments, free access initiatives and portals on the web has reached an unprecedented coverage and will probably keep growing as the

[2] http://www.worldlii.org/worldlii/declaration/

web expands. However, a coherent logical organization of the available material and an effective accessibility in terms of reusability of information, findability, completeness of the legal sources and related information regarding a legal topic has still to come.

Despite the increasing coverage the current scenario in legal information access is in fact in the majority of cases a proliferation of different points of access organized with different criteria for the different legal information collections, countries and publishers. The data are themselves stored and represented in different formats in separate closed and unaccessible databases without any chance to establish an interoperability layer among the different source of information allowing information reuse and correlation

Most of the effort is left to the user to query, collect and integrate the information he needs in a time consuming search activity as a work around to overcome the technical barriers that keep the available legal digital collections separated.

Many initiatives in the legal informatics research community have been promoted to deal with such issues. In particular semantic enhancement of legal information (Casellas, 2010) and national and international initiatives on legal documents standardization. However, despite the relevance of such researches the top down approach they underly has so far been an obstacle to the actual implementation on a large scale of the envisaged solutions as they would require a wide coordination and economic effort of the involved actors to adopt the proposed standards with little immediate benefits.

The current momentum gained by Open Government Data and Linked Open Data initiatives around the world offers a unique opportunity to foster an effective free and open access to legal information according to the linked open data principles joined by the complementary standardization initiatives brought on by the legal information community.

One of the keys to success of the Linked Open Data initiative is that, while based on exactly the same technological stack and principles of the (Legal) Semantic Web community, it proposes a pragmatic bottom up approach to refine and enrich the web resources sketching an immediately viable path towards the incremental realization of the semantic web vision. The first step towards this change of perspective is the exhortation to data owners to *"liberate the data"* or *"raw data now" i.e.* to put the raw data available and accessible in whatever open format, first necessary condition to implement the overlying knowledge organization layers.

By adopting the linked data best practices of publication (Heath, 2011) (first of all on existing databases), we put the basis for a growth free of technical barriers of enhanced content which will constitute the basic building blocks to be later reused to feed more advanced knowledge intensive systems.

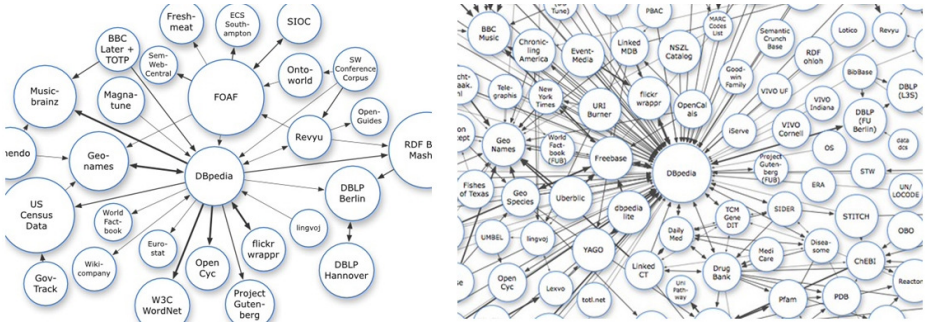

Fig. 1 growth of the Linked Data Cloud from 2007 to 2011[3]

As Fig. 1 witnesses, layered standards allow for an iterative development process that progresses quickly. They help control costs and raise participation in communities where markup to a fine-grained standard may be too costly or burdensome.

As cited in (Holmes, 2011) one of the barriers to accessible law is that

"To a worryingly large extent, statutory law is not practically accessible today, even to the courts whose constitutional duty it is to interpret and enforce it. There are four principal reasons. … First, the majority of legislation is secondary legislation. …Secondly, the volume of legislation has increased very greatly over the last 40 years …Thirdly, on many subjects the legislation cannot be found in a single place, but in a patchwork of primary and secondary legislation. … Fourthly, there is no comprehensive statute law database with hyperlinks which would enable an intelligent person, by using a search engine, to find out all the legislation on a particular topic.
"

The linked data model seem to provide the technological answer to such a question of fragmentation of legal sources by providing the infrastructure to seamlessly collect in a single place document fragments from distributed sources.

4.1. OPEN DATA AND LEGAL STANDARDS

The application of the linked opend data best practices of publication in the legal information domain finds fertile ground in mature initiatives for the definition of standard formats for the identification and representation of legal documents on the web. Here we mention the most relevant.

4.1.1. Legal sources identification

[3] http://richard.cyganiak.de/2007/10/lod/

URN:lex is a proposed Internet standard for legal document identifiers[4]. The URN:LEX namespace aims to facilitate the process of creating URIs for legal sources independent of a document's online availability, location, and access mode. "Sources of law" include any legal document within the domain of legislation (including bills), case law and administrative acts or regulations. This identifier will be used as a way to represent the references (and more generally, any type of relation) among the various sources of law.

A different but compatible approach have been proposed and implemented in the definition of the URI identification schema of British Legislation in one of the most advanced initiatives on publishing legislative data in open format in the portal www.legislation.gov.uk described in (Sheridan 2010). Here a Persistent HTTP URI scheme have been used following the design principles inspired by the object-oriented design and modelling characteristic of the O'Reilly Media PRESTO architecture (Jelliffe, 2008) and conforming to the recommendations of W3C for the design of URIs for the semantic web (Sauermann, 2008);

5.1.2. Legal document structure representation

Initiatives on adoption of XML standards for the representation of legislative document structures and metadata have been brought on both at national and international level in different countries in recent years. They all basically aim to provide a Open XML interchange format for legal and legislative resources thus conforming to requirement of opennes in LOD. To cite the most successful, XML.gov in the U.S., Crown XML Schema in the U.K. provide the most rich and complete datasets made available by governments in open XML. Other initiatives in European countries , like NIR (NormeInRete) standard in Italy or Metalex in the Netherlands have also lead to further development for a panafrican standard (AkomaNtoso) and to the international initiative of Metalex/CEN global interchange standard of legal sources.

4.1.3. Metadata scheme

In legislation.gov.uk a sophisticated metadata model — incorporating FRBR[5], the CEN MetaLex vocabulary, Dublin Core Terms[6], and the Crown Legislation Markup Language — enabling advanced version control and output of descriptive metadata have been adopted providing also all the metadata able to implement a point in time legislative system. In general a very minimal metadata set would consist of a title, an effective date, the name of the issuing body and some sort of permanent identifier (for example conforming to the urn:lex specification). Beyond that, one might add more dates (like date of fficacy, date of publication, dates of stages in the process of drafting and approval), compact descriptions of the legislation and

[4] http://tools.ietf.org/html/draft-spinosa-urn-lex-01
[5] http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records
[6] http://www.dublincore.org/documents/dcmi-terms/

its intended effects, a representation of the legislative process, responsible agencies and organizations, and so on[7].
The more descriptive metadata are attached to the document (or even better, at subparts of it), the more meaningful interconnections among different documents can be established

Despite the desirability of the adoption on a large scale of the above mentioned standards, their spreading tend to be extremely slow unless significant effort and lead is taken by governmental institutions for their adoption.
In this respect one of the advice in the reference guidelines for Putting Government Data online (Berners-Lee, 2009) is that "*There are two philosophies to putting data on the web. The top-down one is to make a corporate or national plan, by getting committees together of all the interested parties, and make a consistent set of terms (ontology) into which everything fits. This in fact takes so long it is often never finished[...]. The other method experience recommends is to do it bottom up. A top-level mandate is extremely valuable, but grass-roots action is essential. Put the data up where it is: join it together later. Do NOT wait until you have a complete schema or ontology to publish data.*"

4.1.4. Legacy Database Schema as source of Metadata

Sticking to such philosophy a fundamental source of metadata are existing legacy relational database structures and schemas. Once they are transformed in an RDF schema and made openly accessible on the web according to the same Linked Data principles, they constitute a fundamental and immediately available KOS (Knowledge Organization System) to the underlying instances of the database. Though not perfectly sound from a Legal Informatics point of view they have the desirable effect to be extremely useful to produce a critical mass motivating a virtuous circle of refinement, standardization and mapping to more rich, refined and shared KOSs.

Indeed there are a number of open source tools for putting relational databases up as Linked Data e.g. *D2RQ platform*[8] and *Triplify*[9] being two.
These are able to analyze the schema of an existing database to create a default mapping file explaining how the database structure actually represents things and their relationships. This mapping file can be used as-is or can be customized and refined. The result of the mapping is a translation of the relational database into and RDF Schema with entries of the DB exposed as RDF instances on the web of data.
In this way the original database structure is opened to become an immediately available source of metadata and interlinked relations

[7] Suggested metadata practices for legislation and regulations http://topics.law.cornell.edu/wiki/lexcraft/suggested_metadata_practices_for_legislation_and_regulations
[8] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/
[9] http://triplify.org/

surrounding the documentary units originally stored in the database also exposed as open data.

Similarly the level of granularity of documents markup can be increased step by step by progressive wider adoption of XML standards instead of simple HTML.

5. A case study on European Legal Sources

The adoption of the linked data principles in the legal domain would allow to reach a level of opening and interconnection of existing legal collection and legal related semantic assets on the model of the famous linked data cloud. Our aim is to start the construction of a "*Legal Data Cloud*" sketched below for some of the European Union relevant legal datasets.

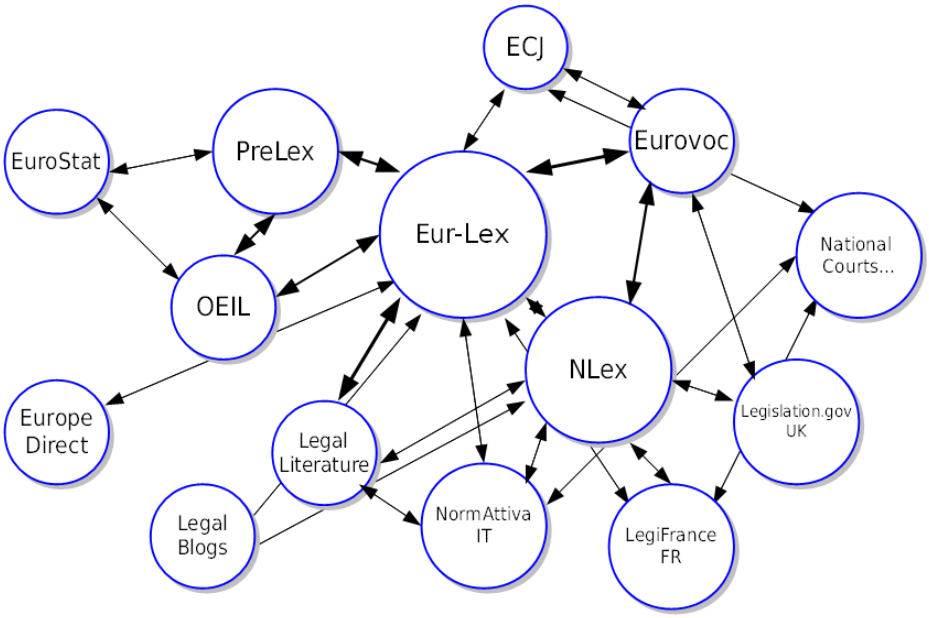

Fig. 2 A possible interconnection of relevant European Legal Datasets in a "Legal Data Cloud".

Here we foresee the interconnection of a number of legal data collection available in Europe of different legal sources (legislation, case law, legal literature) at different level of jurisdiction (European Union level, member states) and at different stages of their lifecycle. From proposal and debate (Pre-Lex monitoring of the decision-making process between institutions, OEIL the legislative observatory of the European Parliament) to publication (Eur-Lex Access to European Law, National Legislation portals like NormAttiva in Italy, LegiFrance in France, legislation.gov.uk in the United Kingdom and so on), to interpretation (legal literature) and comments (e.g. legal blogs) to portals providing informative services to European citizens (EuropeDirect).

The aim is to bootstrap the creation of such a kind of distributed globally interconnected legal information space by adopting the described linked data best practices in the context of existing legal datasets publication.

The interconnection *glue* of such datasets would be constituted at a first stage by interoperability of the original databases schema metadata expressed in standard RDF format, connected by additional mapping RDF statements. Considering that many of such datasets are multilingual, available in parallel translated versions, EuroVoc, the EU's multilingual thesaurus, recently converted in SKOS/RDF format, is another crucial semantic interconnection infrastructure among such data also enabling cross-lingual retrieval. In Fig. 3 a diagram of the entities and properties that can be extracted from the bibliographic notice related to a European Directive is reported.

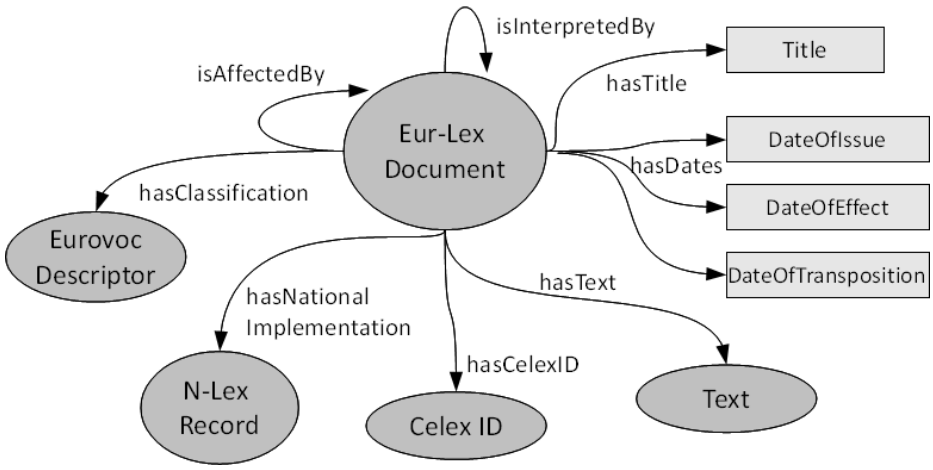

Fig. 3 The metadata scheme underlying the bibliographic notice related to a European Directive as accessible form the Eurlex website http://eur-lex.europa.eu

Such a schema is instantiated by the following RDF/XML serialization of the bibliographic metadata associated to the European Directive 32004L0083.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:eurlex="http://eur-lex.europa.eu/eurlex#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:ev="http://eurovoc.europa.eu/schema#">

        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:hasCelexId>32004L0083</eurlex:hasCelexId>
        </rdf:Description>

        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:hasDateOfIssue>20040429</eurlex:hasDateOfIssue>
        </rdf:Description>
```

```
        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:hasDateOfTransposition>20061010</eurlex:hasDateOfTransposition>
        </rdf:Description>

        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:hasDateOfEffect>20041020</eurlex:hasDateOfEffect>
        </rdf:Description>

        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:hasClassification  rdf:resource="ev:foreignNational"/>
                <eurlex:hasClassification  rdf:resource="ev:rightOfAsylum"/>
                <eurlex:hasClassification  rdf:resource="ev:aidToRefugees"/>
                .........
        </rdf:Description>

        <rdf:Description rdf:about="urn:lex:eu:commission:directive:2004-04-29;2004-83-EC">
                <eurlex:isAffectedBy
                 rdf:resource=" urn:lex:eu:european.court.of.justice:decision:2008-06-16;C-256-
                 08 "
                />
                .........
        </rdf:Description>


      .........
</rdf:RDF>
```

Here we argument the potential benefits of the foreseen open legal data interconnection by reporting some samples from a case study of cross-datasets query from European legal sources, which would be possible once such datasets are made available in RDF.

Queries are here expressed in a SPARQL like syntax. SPARQL is the query language for RDF, analogous to what SQL is for relational databases.

Q1. Assume we want to obtain all the judgements pronounced by the European Court of Justice regarding European Union directives on the subject of right of asylum. For example we can filter from the results only cases regarding Italy.
Accessing the three different datasets of ECJ, Eurlex and Eurovoc thesaurus, exposed in RDF on the web of data, and exploiting the relations established among them, a SPARQL query is able to return only the documents requested.

```
SELECT ?case
FROM <http://eur-lex.europa.eu/eurlex.rdf >
FROM <http://curia.europa.eu/case.rdf>
FROM  <http://eurovoc.europa.eu/eurovoc.rdf">
WHERE { ?directive eurlex:isAffectedBy ?case .
          ?directive eurlex:hasClassification ev:righOfAsylum .
            FILTER(?country='it')
}
```

Q2. Here for example we ask for national decisions on national norms of implementation of European directives on the subject of right of asylum.

Assuming open access to the RDF datasets of national norms and decisions, queried against the eurlex dataset, SPARQL allows to correlate query results of

1. European directives on the requested (Eurovoc) subject;
2. National norms implementing such directives;
3. National decisions on such national regulations;

to obtain in a single shot the desired documents matching the request from different collections.

```
SELECT ?nationalRegulation ?nationalCase
FROM nationalCatalogOfNorms
FROM nationalCatalogOfCases
WHERE { ?eurlexDocument eurlex:hasClassification ev:rightOfAsylum .
    ?nationalRegulation eurlex:isNationaImplementationOf ?eurlexDocument .
    ?nationalCase isCaseDecisionOn ?nationalRegulation .}
}
```

Q3. We can imagine to create a sort of monitoring transparency service, fed by linked legal data, tracking the dates of implementation in national parliaments of recent european directives. This would allow to see for example how different member states are ready in the implementation of European directives, the trends over time, a comparison among different member states etc.
This also could be done via a SPARQL query over the same datasets by selecting the date of publication of national norms *which are implementation* of European directives issued recently (dateOfIssue > data1), and the date of issue of such directives to monitor the delay.

```
SELECT ?dateOfImplementation ?dateOfIssue
FROM nationalCatalogOfNorms
FROM <http://eur-lex.europa.eu/eurlex.rdf >
WHERE { ?dateOfImplementation isDateofPubblicationOf ?nationalRegulation .
        ?nationalRegulation isImplementationOf ?eurLexDocument .
        ?dateOfIssue isDateOfIssueOf ?eurLexDocument .
        FILTER(?dateOfIssue > data1)
}
```

These are just examples of the potentialities of the linked open data infrastructure in the legal domain. Endless possibilities could be immagined by interested parties having free access to such a web of legal data.

What is worth notice is that:

Each of these implementations is nearly costless provided that the underlying legal data cloud is put in place in proper linked open data formats. The cost is only the one of writing a SPARQL query and having it run on one's own web application. This is the idea to give added value to data that are already there

and can be consumed in innovative ways involving the community to provide services hardly immaginable in institutional websites.

One of the technological strength of the linked data approach is that once data are avaialble, access to data is entirely programmatic via SPARQL queries or even more refined Linked Data API (Application Program Interfaces) with little or no user intervention.

The results from cross-datasets queries directly feed a web application. They are themselves open data, typically made avaialble in XML, RDF or other microformats like JSON to be subsequently displayed to the user in any fashion needed. Moreover they are always up to date as the update is delegated to the open data provider.

In this scenario the provision of legal information services e.g. for citizens or professionals the concept of legal databases and portals is overcome by the concept of filtered specialized views, for example by domain or by legal task, on top of a globally interconnected backend of legal data.
The rationale is that, instead of collecting and replicating data to build a website providing a particular service requiring access to legal documents, one reuses raw data exposed by publishers in open format by seamlessly feeding his own web application through queries to the web of legal data via http. Filtering is automatically performed by machine readable programmatic queries in SPARQL and/or conceptual organization stored in machine readable KOS like thesauri or ontologies.

The more the data are interconnected and the transition from documents to data is accomplished through the implementation of granular XML markup and semantic annotation, the more advanced applications can be imagined.

Data from any other available dataset exposed on the linked data cloud can be integrated to add additional related information to one's application. E.g.

Linked data applications can be enriched with other contextual informative data embedded from external sources exposed on the linked data cloud. This can be done automatically by accessing web services allowing string names matching. For example one could integrate informations on the European Court of Justice accessed from the DBPedia dataset (the linked data version of Wikipedia)

Fig. 4 The raw format of the wikipedia page for the European Court of Justice made avaialble in RDF on DBPedia. http://dbpedia.org/resource/European_Court_of_Justice

6. Conclusions

The rapid grow of open government data initiatives around the world, based on the technical recommendations of the linked open data movement, is a unique occasion to foster the publication of legal datasets in open formats towards a full realization of the Free Access to Law principles.

Thanks to a layered standards infrastructure, Linked Data best practices can be transposed as is to legal data publication in order to reach a critical mass of legal information available in linked data format to put in practice the vision of innovative legal semantic web applications and services built on top of a "Legal Data Cloud". Existing legal data standards can be complementary to this process but not constitute a bottleneck in an endless effort of global standardization. Instead, lessons learned from the linked data movement to approach the task of the realization of the semantic web from bottom up should be taken into consideration. Reuse of existing legal metadata scheme from available legal datasets on the web can be a starting point to bootstrap an iterative process of refinement towards a fully interconnected open legal semantic web.

References

Berners-Lee, T. (2006) *Linked Data - Design Issues*, available at http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T. and Berners-Lee, T. (2009) , *Linked Data - The Story So Far,* in: Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems *(IJSWIS)*

Casellas, N. (2010), *Semantic Enhancement of Legal Information… Are We Up for the Challenge?*, available at http://blog.law.cornell.edu/voxpop/2010/02/15/semantic-enhancement-of-legal-informatiom...-are-we-up-for-the-challenge/

Heath, T. and Bizer, C. (2011), *Linked Data: Evolving the Web into a Global Data Space (1st edition).* Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Holmes, N. (2011), *Accessible Law* , available at http://blog.law.cornell.edu/voxpop/2011/02/15/accessible-law/
Sheridan, J. (2010), *legislation.gov.uk*, available at http://blog.law.cornell.edu/voxpop/2010/08/15/legislationgovuk/

Jelliffe, R. (2008), *PRESTO - A WWW Information Architecture for legislation and Public Information systems,* available at http://www.oreillynet.com/xml/blog/2008/02/presto_a_www_information_archi.html

Sauermann, L. And Cyganiak R. (2008), *Cool URIs for the Semantic Web,* W3C Interest Group Note, available at http://www.w3.org/TR/cooluris/

Berners-Lee, T. (2009), *Putting government data online,* available at http://www.w3.org/DesignIssues/GovData.html