



**HAL**  
open science

## Legivoc - connecting law in a changing world

Hughes-Jehan Vibert, Pierre Jouvelot, Benoît Pin

► **To cite this version:**

Hughes-Jehan Vibert, Pierre Jouvelot, Benoît Pin. Legivoc - connecting law in a changing world. Journal of Open Access to Law, 2013, 1 (1), pp 1-19. hal-00937740

**HAL Id: hal-00937740**

**<https://minesparis-psl.hal.science/hal-00937740>**

Submitted on 28 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Legivoc – connecting laws in a changing world**

Hughes-Jehan Vibert\*, Pierre Jouvelot<sup>◦</sup>, Benoît Pin<sup>◦</sup>

\* *Doctor of Law, ITC Project Manager, Ministry of Justice<sup>1</sup>, France*

◦ *Researcher Scientist, MINES ParisTech<sup>2</sup>, France*

**Abstract.** On the Internet, legal information is a sum of national laws. Even in a changing world, law is culturally specific (nation-specific most of the time) and legal concepts only become meaningful when put in the context of a particular legal system. Legivoc aims to be a semantic interface between the subject of law of a State and the other spaces of legal information that it will be led to use. This project will consist of setting up a server of multilingual legal vocabularies from the European Union Member States legal systems, which will be freely available, for other uses via an application programming interface (API).

**Keywords:** law; language; legal internet; legal informatics; thesauri; vocabularies; alignment; member states of the European Union

Legivoc will be a terminology database designed to facilitate the Member States' understanding of EU laws. It is aimed at providing an interoperable terminology system to be used in particular within the framework of legal computer-based projects concerning access to EU Member States laws and the promotion of legal exchanges of information between European networks for legal or judicial cooperation. In an internationalised environment, this system will help rationalise information processing, ultimately allowing easier information research processes as well as a better understanding of foreign laws in a structured, yet legally diverse European scene. Legivoc is a project initiated by the French Ministry of Justice, with the support of the Czech, Spanish, Italian, Finnish and Luxembourgish Ministries of Justice and the European Commission.

---

<sup>1</sup> Available at: <http://www.justice.gouv.fr>.

<sup>2</sup> Available at: <http://www.mines-paristech.eu>.

## 1. Building bridges between laws: why and for whom?

Thanks to the Internet, information can be exchanged much faster within an internationalised, multilingual context. However, as far as legal information requests from a Member State to another go, it is still not possible to meet current needs, as one faces difficulties extending beyond linguistic barriers. Why is it so complicated? Essentially because law is necessarily culture-dependent. Legal terms do not always reflect the same reality, which is noticeable when one compares, for instance, Luxemburg and France together, or Austria and Germany, or even Scotland and the rest of the United Kingdom. Eurovoc<sup>3</sup> is a multilingual thesaurus providing invaluable access to information on the European Union, including legal resources. However its legal content is incidental and does not reflect the diversity of Member States laws. Another tool at one's disposal is IATE<sup>4</sup>, the terminology database of the European Union, which contains as indispensable and invaluable information as Eurovoc and is made out of content collected from international sources; however not all of its sources mention Member States' laws.

Only within the defined scope of national laws, however, does the legal semantic representation of a state or organisation make sense. Thus, while the Eurovoc law domain is the perfect tool when legal information on the European Union is required, it is not quite advanced enough anymore when it comes to finding out information on Member States laws' tinged with either a Romano-Germanic background, a mixture of laws or combining a Common Law basis with national specificities. Yet, there is such a thing as collective intelligence in civil societies<sup>5</sup> and Legivoc's very aim is to appreciate semantic relationships between these different forms of intelligence.

---

<sup>3</sup> Available at: [eurovoc.europa.eu](http://eurovoc.europa.eu).

<sup>4</sup> Available at: [iate.europa.eu](http://iate.europa.eu).

<sup>5</sup> Fazy, James, *De l'intelligence collective des sociétés, cours de législation constitutionnelle, donné par M. James Fazy*, Genève : 1873, Impr. de Vve Blanchard, 450 p.

## 1.1 WHERE CONNECTIONS SHOULD BE MADE

### 1.1.1 *Laws*

At the European level, being able to appreciate national laws is necessary and an inherent part of the building process of evolving EU institutions. This also applies to the broader scale of comparative law research, as establishing semantic relations between legal concepts expressed in different languages and legal systems is a challenging ideal to achieve concretely. European institutions have already been working in this direction, their achievements being actually taken into account on a case-by-case basis in studies providing background material for the drafting process of EU texts.

In order to meet such needs, the Translation Directorate General (DGTrad) of the Court of Justice of the European Union has launched a pilot project called *Comparative multilingual legal vocabulary*, aimed at making available a multilingual documentary tool for indexing and searching the documentary resources of national legal databases from any language, in a uniformed way. This work has a very considerable ambit as it addresses a huge diversity of legal issues and will entail considerable costs.

On July 1<sup>st</sup>, 2013, the European Union constituted 28 Member Countries and 24 official languages. According to a 2010 report from the Translation Directorate General of the Court of Justice of the European Union, it would take 34 legal/linguistic experts working for 20 weeks to process 400 concepts<sup>6</sup>. Of note, in comparison, the UTU thesaurus used by Belgian judicial authorities comprises 8,000 terms. A calculation determined that it would take DGTrad over 8 years to fully process only the terms specific to Belgian law (expressed into 3 official languages) - an absurdly long time. Let us take this line of reasoning further and consider a project initiated in Lithuania which includes

---

<sup>6</sup> Court of Justice of the European Union - Directorate General for Translation, *Vocabulaire juridique multilingue comparé: Bilan sur l'étude de faisabilité* (Caroline Reichling), 30 juin 2010, [http://curia.europa.eu/trad/bilan\\_vocabulaire\\_juridique\\_multilingue\\_compare.pdf](http://curia.europa.eu/trad/bilan_vocabulaire_juridique_multilingue_compare.pdf) (only in French).

40,000 terms pertaining to law and political sciences translated into French<sup>7</sup>: it turns into a 40-year long utopian dream, likely to take longer when one considers the constant expansion of the European Union and the necessary semantic connections that are likely to be required for 30 European States.

Legivoc aims to support this complex and ambitious approach and to be used in turn in support of future additional developments, since, in our opinion, the work achieved so far by DGTrad is as close to perfection as possible. Indeed, manually comparing national legal concepts with a view of establishing connections between them is the most appropriate approach, the least likely to result in inaccuracies potentially induced by automatic/simple translation. Besides, the small number of topics already covered by DGTrad will not need to be re-examined. Instead, they will provide a reliable basis for us, regardless of occasional and unavoidable mistakes or lacunas.

However, Legivoc will mainly depend on the contributions of Member States of the European Union and, when appropriate, on additional work implemented to create national legal thesauri. Legivoc adopts an almost simplistic approach, i.e. bringing elements of national laws together according to each of their specific organisations and official language(s). Further to a request for contributions from Member States of the European Union, about fifteen national legal vocabularies will soon be made available to allow conceptual connecting processes and for their application at a national level and for comparative law studies.

### 1.1.2 Languages

In Lithuanian, *sudaryti rašytinę sutartį* refers to the notion of *written contract*. In order to fully grasp the extent of such a notion, one needs to understand its meaning via another language. However, translating the words will not be enough. One may also wonder what the notion of contract actually implies in Lithuanian law, and what *written* would mean under that law, thus implicitly questioning the concept of evidence. Foreign authors can make word assumptions, but as far as the in depth meaning of the concept is concerned, a

---

<sup>7</sup> *Lietuvių - Prancūzų- teisės ir politikos žodynas*, online, <http://www.zodynastp.lt>

misunderstanding could occur. The key is to appreciate all the concepts interconnecting within one same notion, whose understanding requires a firm knowledge of the legal system they refer to.

Literal translation would be an easy solution, which would have the unfortunate consequence of paving the way for big misunderstandings and misinterpretations. In his thesis, lawyer-linguist Guy Mazet<sup>8</sup> published the results of 30 years of research and used terms of French and Italian laws as examples: the terms *amministratore* and *administrateur* for instance seem to be equivalent. However, the former is used to describe all directors, regardless of a company structure, whereas the latter only applies to *sociétés anonymes* (anonymous society/company, or “public limited companies” in the Common law countries). However, does the concept of *société anonyme* exist outside of France? If so, how about *administrateur* in Belgian or Luxembourgish laws, where linguistics aspects are not an issue, law texts being written in either French only or French and other official languages. Could computer-assisted translation tools take into consideration all the subtle details from each Member States legal corpus which necessarily originate from unique cultural, historical and political backgrounds? Most probably not, or at least not without some prior preparatory work, and this is precisely where Legivoc steps in.

Lawyer-linguist Guy Mazet wrote in his thesis that “in order to build a bilingual tool granting access to legal terminology databases, translation will not be the only issue. Interlinguistic equivalence relationships are not sufficient to convey the modalities of judicial systems connecting to one another. They shall be kept as such under the right circumstances, but one must search for the depth of the degree of relationship intrinsically binding two legal systems: that degree is called functional equivalence”<sup>9</sup>. Considering there are more than 28 legal systems (regardless of national particularities) where 24 official languages are used (regardless of the national languages that are not part of the European Union official languages), Legivoc follows Mazet’s views, massively, but equally strictly. However, it is not in light of the

---

<sup>8</sup> Mazet, Guy, *Jurilinguistique et informatique juridique*, 2001, IRETIJ, Montpellier I University.

<sup>9</sup> *Ibid* p. 18.

worthwhile nature of this project and the collaborative approach that will be adopted.

## 1.2 IN VARIETATE CONCORDIA<sup>10</sup>

*Languages, laws, and connections between laws* is what Legivoc is about and summarises the project. We are seeking to establish functional equivalents of legal terminology and, for the purposes of this project, the linguistic aspects of the terminology are not afforded much importance. Whilst translation has its place, it is merely considered a tool possibly helping or to the contrary hindering the legal understanding of specific legal terms, in light of such variables such as the quality of the available translation and the background knowledge required for jurisdictionally specific terms. As is the case for each of the Member States of the European Union, legislation in each country's designated official language(s) will be the reference point on Legivoc.

This approach confers great flexibility. Almost paradoxically, it will allow us in future, to add an unlimited number of translations to the database if required. In effect, French, Maltese or German law translated into English, Spanish or Greek (or into Arabic, Russian, Chinese etc...) lacks legitimacy. This is because only the official language of the considered legislation provides the required value perspective and reference points necessary to understand the true meaning and implications of the law and legal terminology involved. A simple translation can only be used as an indication of the meaning of the law. Achieving a term's proper legal translation will only be made possible by establishing functional equivalences between the national legal concepts of the source legal system and the state requiring the translation.

Accordingly, our approach will be a practical and rational one. In a first instance, each national vocabulary within legivoc will be available in all the official languages of the European Union. The legal vocabulary provided by a

---

<sup>10</sup> *United in diversity* is the motto of the European union, "it signifies how Europeans have come together, in the form of the EU, to work for peace and prosperity, while at the same time being enriched by the continent's many different cultures, traditions and languages". Available at: <http://europa.eu/about-eu/basic-information/symbols/motto/>.

member State will be the reference point, its literal translation will be given as a simple indication. The translations will be generated by automated translation tools (MT@EC, the Machine Translation Service of the European commission for the European Union official languages<sup>11</sup>) and users will have the possibility to suggest better alternatives.

### Belgian part of legivoc (example)

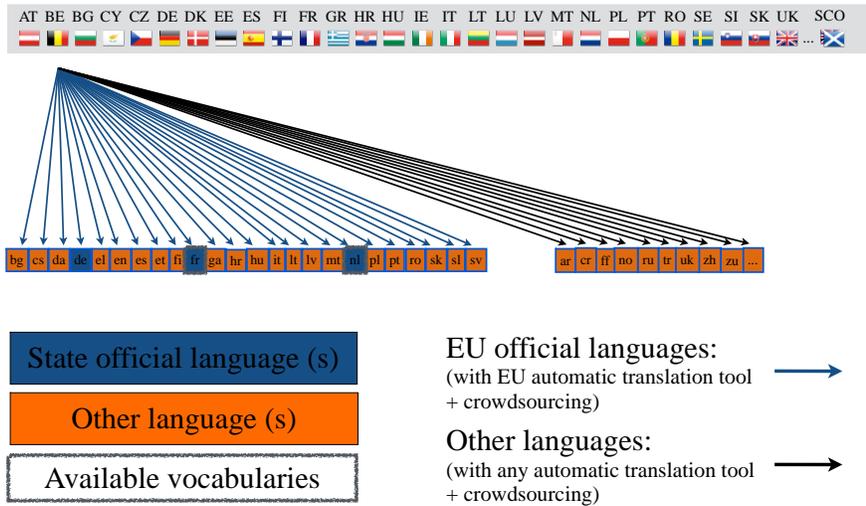


Fig. 1. Legivoc approach for each state with a focus on Belgium whose legal vocabulary is to be available on Legivoc in 2 of its 3 official languages

The graph below highlights some of the information we have at our disposal,

<sup>11</sup> Available at: [http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action\\_en.htm](http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm).

*i.e.* UTU, the Belgian legal thesaurus for Belgium, which is available in 2 of its 3 official languages. As circled, the concept of *demande additionnelle* or *aanvullende eis* (*i.e.*, *additional claim*, or rather *additional demand*) corresponds to one single concept, which is specific to Belgian law and marked by the number 20. It goes without saying that translating Concept number 20 into any language will be possible, translations being performed following the Belgian thesaurus' hierarchical order 1 > 2 > 11 > 13 > 20. Web addresses and references will define the law, not words.



DROIT JUDICIAIRE	1	GERECHTELIJK RECHT
DROIT JUDICIAIRE - PRINCIPES GÉNÉRAUX	2	GERECHTELIJK RECHT- ALGEMENE BEGINSELEN
Principes (droit judiciaire - principes généraux)	3	Beginselen (gerechtelijk recht - algemene beginselen)
Généralités	4	Algemeen
Droit commun de la procédure	5	Gemeen procesrecht
Droit transitoire	6	Overgangsrecht
Déni de justice	7	Rechtsweigerig
Prononciation par voie de disposition générale et réglementaire	8	Uitspraak algemene regel
Lois interprétatives	9	Uitleggingswetten
Compétence	10	Bevoegdheid
Notions	11	Begrippen
Généralités	12	Algemeen
Demande	13	Vordering (eis)
Généralités	14	Algemeen
Demande	15	Eis
Demande incidente	16	Tusseneis
Demande reconventionnelle	17	Tegeneis
Demande en intervention	18	Tussenkomst
Demande nouvelle	19	Nieuwe eis
Demande additionnelle	20	Aanvullende eis
Jugement - arrêt	21	Vonnis - arrest
Généralités	22	Algemeen
...		
Politique étrangère et de sécurité commune	8381	Gemeenschappelijk buitenlands en veiligheidsbeleid
Énergie - Droit européen	8382	Energie - Europees recht
Protection civile - Droit européen	8383	Civiele bescherming - Europees recht

Fig. 2. Extract from Belgian thesaurus UTU in 2 of the 3 official languages

Although it does not encompass any creative process, this part is crucial: Legivoc only presents the EU Member States' future contribution through their ministries, academic/professional bodies or legal resources centres. This initiation phase is merely a way to bring entities closer. In September 2013, we will have acquired about 15 national legal vocabularies out of 28 Member States, most of which will be available in an easy to use format, some structured and organised, some appearing as a simple list of words, some translated into other languages, some comprehensive, and some quite incomplete.

With its multifaceted dimension – several nations, laws, languages, forms and approaches – Legivoc is quite simply revealing of a diversity that shall be preserved, including during the rationalising phase of experimenting semantic correlation of legal concepts. Because of its possible and potential future uses, we insist on keeping its structure as flexible as possible, both in its organisation and the provision of national legal concepts.

## **2. How laws connect to each other on Legivoc**

To date, we have about 15 national vocabularies in store. Of importance will be to ensure the preservation of their original display and functioning mode while guaranteeing their availability in a single format. The database will give access to the vocabularies we have collected in at least one official language and will provide available translations into other languages at a later stage. This aspect indubitably suffices to make the project worth launching. This first phase is indispensable before the terminology can have other uses, and will not need to be implemented anymore afterwards, as other participants may be in charge of improving the handling of vocabularies (depending on each Member State's licence agreement).

As stated, the work done by DGTrad of the Court of Justice of the European Union will be available to Legivoc. It mainly consists of researching conceptual equivalents, therefore establishing interconnections between national legal concepts. As the on-going project requires around 30 lawyer

linguists working over several years, while ideal in terms of quality it will also be costly by reason of the human factor involved. We would like to keep this human-oriented approach in a collaborative process through the promotion of interactions between the various contributors to Legivoc by promoting areas of exchange.

We are in charge for the technical administration of two other projects that may support the Legivoc project – Legivoc having been added to them before. The first one is Legicoop, the website of the Network for Legislative Cooperation between the Ministries of Justice of the European Union (NLCEU), and the second Babellex, a virtual marketplace for job-related offers and requests to/from legal translation and interpretation professionals, which also features legal resources for translators and interpreters.

## 2.1 CONNECTIONS TO ON-GOING MONITORED PROJECTS

### 2.1.1 *Legivoc and the Network for Legislative Cooperation between the Ministries of Justice of the European Union*

The NLCEU<sup>12</sup> website's three-dimensional approach is characterized by cultural, technical and also human factors. As a website dedicated to information exchange, it involved participatory processes. One of the key issues that needed to be addressed when it was designed was how to feed continued development impetus during the website's first three years of implementation to avoid its stagnation.

As of July 1st, 2013, there are 24 official languages and 28 legal systems in the European Union (even more if one takes into account regions, Länder and nations like Scotland). The European motto *in varietate concordia* – United in diversity – summarizes the context. The NLCEU website's doesn't obscure the cultural dimension of the project, by promoting exchanges, actions and

---

<sup>12</sup> See the *Resolution of the Council and of the Representatives of the Governments of the Member States, meeting within the Council, on the establishment of a Network for legislative cooperation between the Ministries of Justice of the European Union*, Official Journal C 326, 20/12/2008 P. 0001 – 0002, online, Available at: [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:42008X1220\(01\):EN:HTML](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:42008X1220(01):EN:HTML).

interactions between individuals. Since its launch, 29 correspondents have been designated by their respective ministries (the United Kingdom having two, one for England and Wales, another for Scotland). This tool's aim is to establish connections between diverse legal and judicial systems in order to promote mutual knowledge and understanding on legal issues common to its members.

As it is a virtual tool designed for all members, the IT skills required to contribute were defined based on the lowest common denominator. Keeping the *modus operandi* previously used to exchange information by maintaining some processes of the first NCLEU's mailing list seemed most relevant. Email thus remained the preferred medium for exchanging information: whenever a request or an answer is posted on the Legicoop website, all correspondents receive an email with a link to read the request and/or edit an answer directly on the website. This way, exchanges, communication between correspondents and content supply happen simultaneously and unequivocally on NCLEU's website Legicoop.

The website was thus designed as an easy-to-use tool with complex ends, multi-faceted in its languages, legal systems and even uses, adapted to computers, mobile phones and tablets (used in some ministrie

# legicoop

Communication language (same as before)

National language

**Title of the request**

**Titolo della richiesta**

**Descriptor**

**Descrittore**

1 2 3 4 ...

1 2 3 4 ...

**Request**

**Richiesta**

**(Auto)Reply for the Italian law**

**Rispondi per la legge italiana**

Attached documents (translated or not)

Attached links

legivoc

required

Fig. 3. NLCEU website's schematic view

The languages used by national correspondents on the NLCEU website to exchange have to be understandable by the requesting country(ies), often English; however information about each participant's language must also be provided (see above from the Italian correspondent: question title, some keywords). Legivoc shall entail all this information. Each national correspondent will therefore have a reference vocabulary on his/her law at his/her disposal, thereby allowing for quality improvements in the information flow.

Following a pretty straightforward process and further to an issue raised in a State request, other States will provide an answer based on their specific

judicial system. The graph below illustrates this simple process. Request by request, connections between legal concepts will be made on an issue common to Member States.

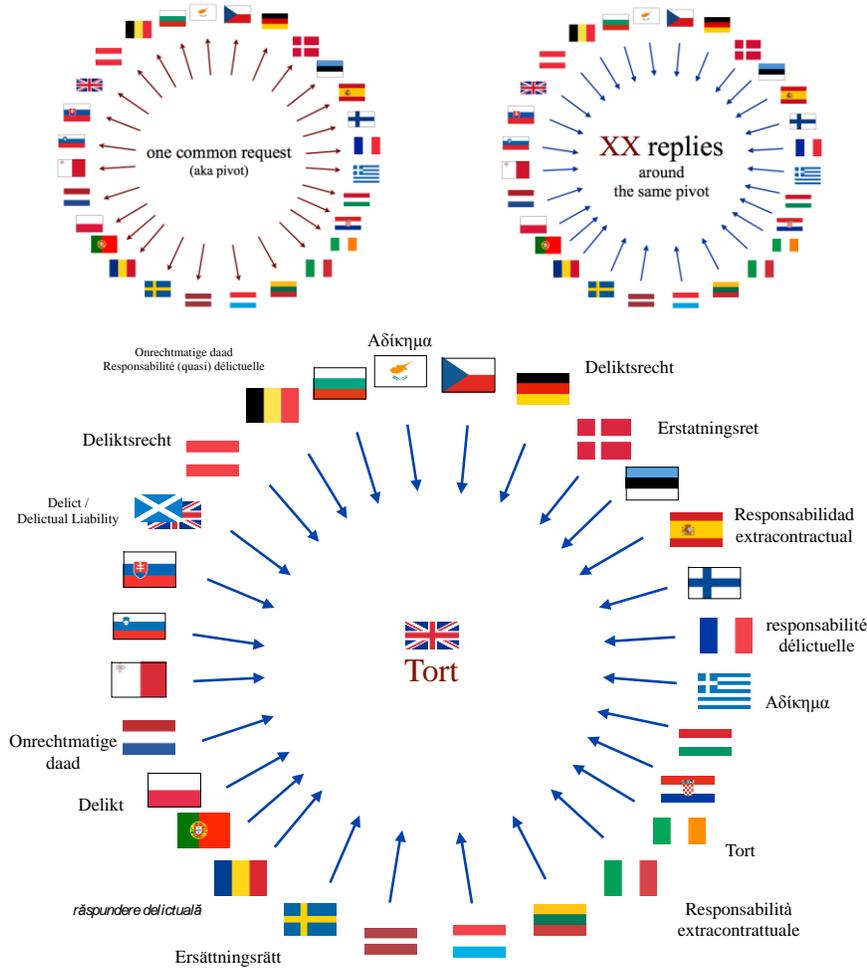


Fig. 4. The pivots are legal points, not a language (juridically approximate example)

Establishing relevant connections between national legal terms will obviously take time. Only increasing the number of exchanges will allow us to hope to achieve relevance, which is why Legicoop, a tool funded by the European Union along with 5 member states' justice ministries will, upon request, be put at the disposal of any other network promoting information exchanges, Legicoop's use bringing Legivoc into play. All connected elements will in turn be added to the Legivoc's server, keywords being the sole reference, within the framework of exchanges and according to modalities further developed below.

### 2.1.2 *Legivoc and Babellex*

The ongoing Babellex project is being carried out jointly with Legivoc. It is co-funded by France, Estonia, Austria, Germany and the European Union and coordinated by the French Ministry of Justice. This project includes creating an online space where legal translators can be asked for help within the framework of any legal proceedings requiring translation or interpretation services<sup>13</sup>. It will also aim to build up an online library with legal translation resources collected from past translations, including judicial documents. This second phase will enhance Legivoc impact, as a tool providing translation professionals with online help on legal concepts<sup>14</sup>.

Crowdsourcing work, that is an approach focusing on the human factor, will be carried out to establish connections between national concepts, although this time around for translation purposes only. Launched on the European area, this project has potential applications to all written languages. Through this approach, some translations of concepts from Member States of the European Union will be made available in languages from other territories.

Such virtual spaces dedicated to connecting legal concepts or promoting accurate translations are invaluable. Experts, lawyer-linguists and professional legal translators will be working on this project, as they are the ones

---

<sup>13</sup> See *Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings*, online Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32010L0064:EN:NOT>.

<sup>14</sup> States will decide upon allowing use and re-use of translations, texts and concepts.

possessing the most extensive knowledge of their own judicial system. Such talents form the core of this connecting process and their work will result in establishing the credibility of the project. Moreover, we can also foresee that Legivoc will be receiving contributions from a much larger audience. To this end, Legivoc will probably be accessible as an educational tool allowing the general audience to freely contribute by suggesting connections (between languages and laws) and it will be our job to appreciate their relevance. Legivoc API will hereby act as a gateway to such contributions subjected to future thorough analysis.

## 2.2 CREATING CONNECTIONS THROUGH OTHER NETWORKS: THE LEGIVOC WEB INFORMATION SYSTEM

Managing the legal textual vocabularies of dozens of countries calls for the design and implementation of a powerful yet highly flexible information system, tentatively called LeWIS (Legivoc Web Information System). One of our key design goals is to avoid requiring modifications to existing data exchange procedures among partners while providing a new and general development pathway to anyone interested in adopting what we see as “best practices”.

LeWIS target architecture is illustrated Figure 5. It is structured around a main database that stores legislative thesauri and vocabularies provided by participating countries. Given the current lack of normative specifications regarding legislative data, juridical bodies use different formats for storing their information and adopt different semantic definitions for them. To adapt seamlessly to this state of affairs, we plan to develop *transcoders* to translate various existing formats towards our own Legivoc internal format (.lgy files). The specification of the LGV Domain Specific Language (DSL), to be implemented as an XML derivation quite probably following some variant of Simple Knowledge Organization System (SKOS) Core<sup>15</sup>, is currently under way. Once this language definition is completed, our goal is to have LeWIS accept as input any file written in variations of XML (.xml), Excel (.xls), text

---

<sup>15</sup> Available at: [www.w3.org/2004/02/skos](http://www.w3.org/2004/02/skos).

(.txt) and, ultimately Legivoc (.lgv) files, if partners hopefully adopt this format. Given the constant updating of law vocabularies all over Europe, automation of these transcoding processes will be a key productivity issue. Obviously, the reverse process, i.e., enabling the transcoding of .lgv files into various existing output formats, will also be provided, to allow the use of external tools.

Since offering multimodal accesses to vocabularies is at the core of Legivoc, we are considering developing two main interfaces to this database: a Web interface and an Application Programming Interface (API). These two interfaces will be accessible via a Web-enabled server, ensuring widespread availability of LeWIS resources to a variety of clients, from desktops to mobiles. The first interface is browser-based. When completed, the site [legivoc.eu](http://legivoc.eu) will allow registered users to access the judicial, legislative and law concepts, as provided by partners in each, together with its translation in all official languages of the European Union. Such pages (one concept per HTML page) will also contain additional information, when available, regarding related concepts, synonyms, homonyms... and their translations; the representation of such relationships will take advantage of SKOS linking features. To speed-up the development process of this user-centric interface and build as soon as possible a community around it, we are investigating the possibility of using one of the few various existing thesaurus management systems.

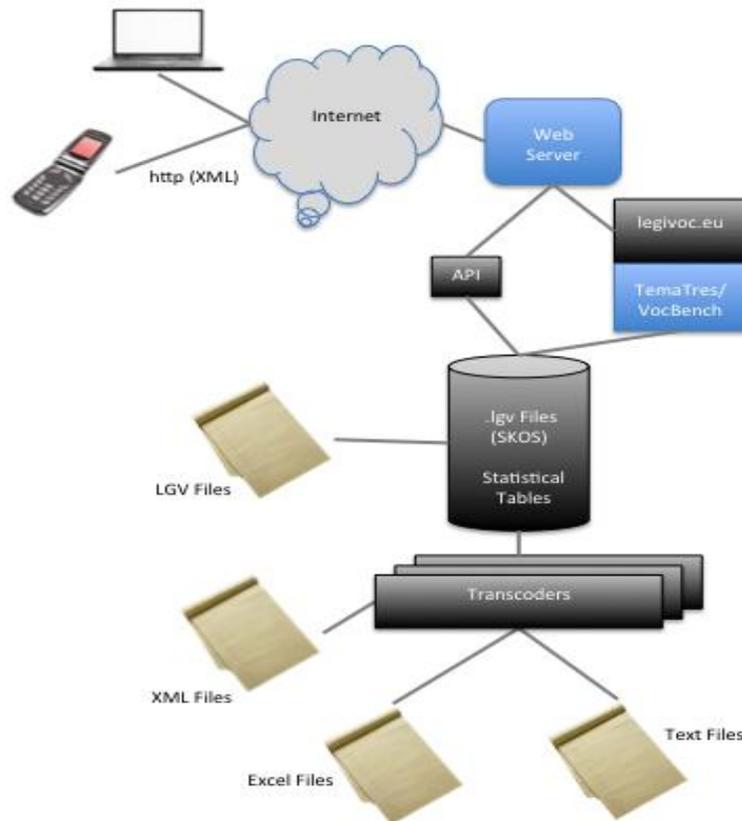


Fig. 5. In this architecture overview, boxes in black are LeWIS-specific

One of them is TemaTres<sup>16</sup>, based on the PHP programming language and which uses MySQL as its database management system (DBMS). This same DBMS is at the core of Java-based VocBench<sup>17</sup>. Both systems have advantages

<sup>16</sup> Available at: <http://www.vocabularyserver.com>.

<sup>17</sup> Available at: <http://aims.fao.org/tools/vocbench-2>.

and drawbacks: for instance, TemaTres is an open-source project and is linked to thriving user and developer communities, while VocBench profits from a more mature implementation language, namely Java. Whether we use for Legivoc one of the above mentioned systems or decide to design a dedicated system, we will make sure that we are able to use our framework of choice, namely SKOS, to encode and store .lgv files.

The second interface will enable third-party programs to access Legivoc data via the Internet through a dedicated Application Programming Interface, to be designed in the very near future. Applying the well-established Representational State Transfer (RESTful) design model<sup>18</sup>, it will allow direct access to the LeWIS database, using various interrogation and update modes<sup>19</sup>. Even though thesaurus management systems already provide some thesaurus access facilities through their own APIs, we are contemplating developing the LeWIS API independently from this existing basis. One reason for this choice is that, even though rights management and security issues are not central to our project for the time being, LeWIS intends to ensure from the start that a secure access service can, ultimately, be derived from its API. PHP, used to implement TemaTres, may not be the best tool to reach that goal; also, TemaTres is not a full-fledged framework, which makes extending its service set not as straightforward as might be hoped. VocBench has more assets on this particular issue.

One key goal of our new API is to enable LeWIS to unobtrusively keep track of user activities (in a detailed yet anonymous manner), and deduce semantically rich information from such accesses. A set of dedicated database tables within the DBMS will be constantly updated accordingly, for statistical, “Big Data” analysis purposes. A long-term goal of this access data gathering process is to compute statistical correlations between various access and usage modes, which should help establish semantic relationships between legislative concepts, both within a given country and across countries. For instance, the set of terms searched by a particular user in a given session might suggest that

---

<sup>18</sup> Available at: <http://www.ibm.com/developerworks/webservices/library/ws-restful>.

<sup>19</sup> Only well-identified submitters have modification rights on key LeWIS data.

the corresponding terms are semantically related; this is even more true if information regarding the users' knowledge level (e.g., jurist, neophyte, official translator) is available.

During the development of LeWIS, we intend to loosely follow the Research-Action approach pioneered by Kurt Lewin, tightly coupling the design, implementation and experimentation phases with expert jurists along the way. A dedicated platform (based either on existing frameworks such as TemaTres or Vocbench, or developed in-house), well-accepted tools (SKOS) and agile programming techniques based upon versatile dynamic languages such as Python, for the development of workflow-related processes, will be the enabling technologies.

### **3. Conclusion**

Legivoc provides both guidelines and a set of solutions originating mostly in the crowdsourcing work of practising lawyer-linguists, on the basis that connecting laws from different legal systems via online tools dedicated to transnational legal resources, can only happen by successfully addressing and discussing concrete legal cases and issues. It is of note that, as far as Legivoc is concerned, a human-centric approach is favoured to encourage rigour. It is hoped that it will allow build foundations for other types of solutions, including automatic processing that will facilitate the highlighting of homonymic and implied terms, mistranslations as well as other tricky semantic-related issues encountered when comparing different laws and languages. The vocabularies and their connexions will soon be accessible in a single SKOS format, simply, directly and freely, wherever possible.