

Business Requirements for Legal Knowledge Graph: the LYNX Platform

Jorge Gonzalez-Conejero^{a,1}, Pompeu Casanovas^{b, a} and Emma Teodoro^a

^a*Institute of Law and Technology – Universitat Autònoma de Barcelona*

^b*Data to Decisions Cooperative Research Centre, La Trobe Law School, La Trobe University, Melbourne*

Abstract. European small and medium-size enterprises (SMEs) and large corporations face multiple constraints to engage in trade abroad and to localize their products and services to other countries, mainly as a consequence of legal and language barriers. This is one of the main consequences of the multiple differences across Europe, which is fragmented into legal silos and into more than 20 linguistic islands. LYNX H2020 project will provide more effective ways of accessing huge amount of digital regulatory compliance documents, including legislation, case law, standards, industry norms and best practices. In particular, the LYNX project envisages an ecosystem of smart cloud services to better manage compliance documents, based on a Legal Knowledge Graph (LKG) which integrates and links heterogeneous compliance data sources. This ecosystem will enable smart search, smart assistance and smart referencing of case law, as well as Artificial Intelligence technologies and machine translation of regulatory compliance documents. An initial step in the development of the LYNX platform is the collection of business requirements from end-users and relevant stakeholders. Therefore, this work introduces the techniques used for the gathering of business requirements from end-users and stakeholders and a list of prioritized business requirements collected through qualitative and quantitative techniques.

Keywords. Compliance, legal knowledge graph, smart cloud services, business requirements

1. Introduction

The European market is fragmented into legal silos and into more than 20 linguistic islands, which constitutes a competitive disadvantage for SMEs and companies in general. Therefore, dealing with legal and regulatory compliance data is a cumbersome task usually delegated to law and consultancy firms, who have to obtain documents from several data sources, published by various institutions according to different criteria and formats by various institutions.

The main objective of LYNX is to create an ecosystem of smart cloud services to better manage compliance, based on a legal knowledge graph (LKG) which integrates and links heterogeneous compliance data sources including legislation, case law, standards and other aspects. This cloud of services integrated in the Lynx platform will

¹ Corresponding Author, Jorge Gonzalez-Conejero, Institute of Law and Technology, Universitat Autònoma de Barcelona, Campus de la UAB, Cerdanyola del Valles (08190) Spain; E-mail: jorge.gonzalez.conejero@uab.cat.

provide mass-customized regulatory information (including legislation, regulations, and policies) to European businesses.

The aim of this work is to collect all the business requirements provided by end-users and relevant stakeholders (SMEs, Large Enterprises, Law firms, among others). Quantitative and qualitative techniques have been used in order to gather and prioritize each of the business requirements identified.

This work is structured as follows: Section 2 lists some legal and business requirements for compliance works and European projects and briefs Legal Compliance by Design (LCbD) and Legal Compliance through Design concepts (LCtD); Section 3 describes the process used for the elicitation of the business requirements; and finally, Section 4 points out the results obtained from the Knowledge acquisition process carried out in the previous Section.

2. Legal Compliance

Legal and business requirements for compliance (especially for compliance by design) have attracted much attention [1]. Previous EU projects—especially COMPAS², OPENLAWS³, EU Cases, MIREL⁴, and BO-ECLI—have developed conceptual toolkits. Moreover, the Workshop on Requirements Engineering and Law (RELAW)⁵, has been running from ten years now, led by specialized researchers such as Sepideh Ghanavati and Guido Boella.

In a previous edition of the LYNX Workshop on Legal and Regulatory Compliance (TERECOM), we presented some preliminary results from the survey we are carrying out [2], after examining 280 works on Compliance by Design in the past fifteen years. After examination of the state of the art, we suggested the concept of Legal Compliance through Design (LCtD) to complement LCbD by recognizing the role of social, political, and economic conditions (as pre-conditions) and governance and ethical requirements (as constraints) when designing legal compliance, encompassing norms and principles that require a balancing of competing rights, obligations or policies. Conditions for legal compliance are broader and more entangled than for regulatory compliance, as legal conditions can be described by means of rules, but rules alone do not play out the stakeholders' rights, duties, and legal effects of their behavior.

We focused on the definition of legal (not only documentary) sources to select and define requirements. Compliance through Design (CtD) explicitly encompasses the social and institutional aspects that are not explicitly included by the regular way of approaching this subject (i.e. legal interpretation processes —beyond the conversations between experts and computer scientists—, institutionalization, the interface between modelling and coordination, and the relation between citizens, consumers, and the law).

This is coherent with Motta's assertion about the interdisciplinarity of descriptive empirical approaches [3], and with the need to consider software requirements as prescriptive statements.

² COMPAS: https://cordis.europa.eu/project/rcn/85292_en.html%20

³ OPENLAWS: <https://info.openlaws.com/openlaws-eu/>

⁴ MIREL: <http://www.mirelproject.eu/>

⁵ IEEE Requirements Engineering and Law (RELAW) Conference: <https://ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1002649>

Thus, the results summarized in Table 2 and Table 3 could be reframed into a general classification of legal sources, properties, and entity relations, respecting the autonomy and decision-making capacity of lawyers, rulers, administrators, companies, business-holders, and lay-people. This is compatible with the LYNX approach as well.

3. Knowledge Acquisition Process

The Knowledge Acquisition Process (KAP) performed—following the Value Proposition Canvas⁶—consists of quantitative (survey) and qualitative (interviews and focus groups) techniques. These techniques are applied to relevant end-users and stakeholders outside the LYNX project use cases. The profiles of these relevant end-users and stakeholders are listed in Section 3.1.

The KAP is specifically devised to provide the LYNX consortium with the outmost business requirements in three different subjects when dealing with digital regulatory compliance documents:

- **Strategy** for the search, analysis, processing, monitoring and handling of digital regulatory compliance documents. It includes workflows and/or strategies for the analysis, process and management of digital regulatory compliant documents.
- **Pains** when dealing with digital regulatory compliant documents. This subject is focused on collecting anything that could annoy customers/entities before, during and after dealing with digital regulatory compliant documents.
- **Gains** when dealing with digital regulatory compliant documents. This subject is focused on collecting outcomes and benefits from customers/entities when dealing with the analysis, processing and managing of regulatory compliant documents.

Therefore, Section 3.1 defines the role, description and requirements of the end-users and stakeholders for their participation in the KAP phase; Section 3.2 describes the quantitative (survey) stage performed within the KAP; and in Section 3.3 introduces the qualitative (interviews and focus groups) stage of the KAP process designed for the LYNX project.

3.1. Targeted end-users and stakeholders

The end-users and stakeholders targeted for the knowledge acquisition process (surveys, interviews and focus groups) are described in Table 1. In this table the description and requirements are listed. The list of requirements is not exhaustive; any institution with relevant knowledge or know-how for the LYNX project is suitable to participate in this phase.

Table 1. Targeted end-users and stakeholders

End-user/stakeholder	Description	Requirements
----------------------	-------------	--------------

⁶ Value Proposition Canvas was introduced by Alex Osterwalder: <https://strategyzer.com/canvas/value-proposition-canvas>

Consultancy firms	Enterprise that provides advice to another entity	Domains: <ul style="list-style-type: none"> • Big data • Legal • Semantics • Internationalization
Legal Advisor	Law firm or lawyer	Domains: <ul style="list-style-type: none"> • Legal • Experience with the regulatory compliance scenario
SMEs	<ul style="list-style-type: none"> • Less than 250 staff headcounts • Less or equal of 50M euros turnover • Or less or equal of 43M euros balance sheet total 	Domains: <ul style="list-style-type: none"> • Enterprise that develops software related to one of the following topics: big data, semantics, natural language processing. This list is not exhaustive. • Internationalized enterprise • Enterprise in process of internationalization
LEs	<ul style="list-style-type: none"> • More than 250 staff headcounts • More of 50M euros turnover • Or more of 43M euros balance sheet total 	Domains: <ul style="list-style-type: none"> • Enterprise that develops software related to one of the following topics: big data, semantics, natural language processing. This list is not exhaustive. • Internationalized enterprise • Enterprise in process of internationalization
Public or private agencies	Public or private agencies in the internationalization domain and professionally involved	Domains: <ul style="list-style-type: none"> • Public or private agency that helps companies in the internationalization process

3.2. Survey

The LYNX survey design process relies on two main pillars: (i) the identification of relevant end-users and stakeholders and the requirements that make them suitable for the LYNX scenario (Table 1); and (ii) the Value Proposition Canvas for the design of the questionnaire.

The Value Proposition Canvas helps to design products and services that end-users and stakeholders really want because it allows to focus on what matters most to them. Jobs to be done by end-users and stakeholders is one of the main inputs since jobs describe the things that end-users and stakeholders are trying to get done in their work or in their life. A job could be the tasks they are trying to perform and complete, the problems they are trying to solve, or the needs they are trying to satisfy.

What are the stepping-stones? What are the contexts? How do the activities change depending on these contexts? What functional problems are end-users and stakeholders trying to solve? These are some of the questions involved in the Value Proposition

Canvas. As a result, Figure 1 depicts the survey design scheme developed for the LYNX survey.

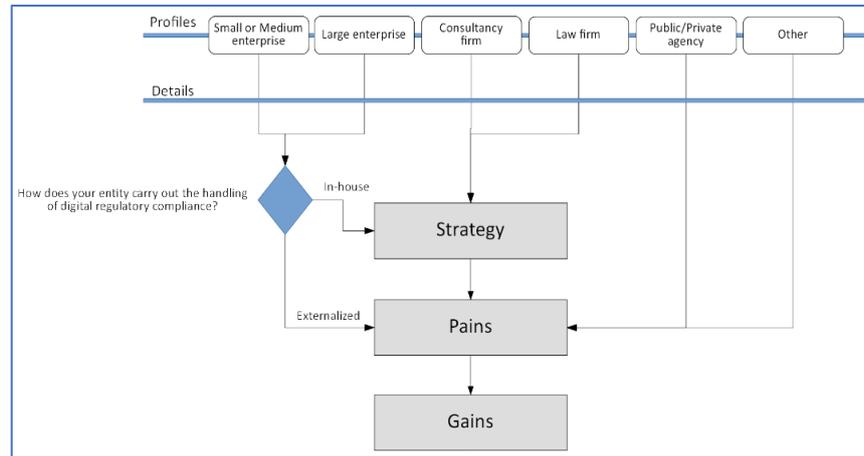


Figure 1. Survey design scheme.

The final questionnaire obtained from the survey design process is published in [4]. It also contains the Electronic Consent; the organization profiles; the strategy for the search, analysis, processing, monitoring and handling of digital regulatory compliance documents; pains and gains. A total of 120 e-mails were sent out with invitations to answer the questionnaire. As a result, 15 of the contacted organizations answered the survey. The distribution by country and organization profile is listed in [4].

3.3. Interviews and Focus Groups

A “Qualitative Interview” is a method of collecting rich and detailed information about how individuals experience, understand and explain certain events or particular topics [5]–[8]. Interviews are “semi-structured” because the interviewer has a list of questions or key points to be covered during the interview and works through them in a methodical manner. Similar questions are asked to each interviewee, although supplementary questions could be asked as appropriate. In general, questions are worded so that responses are open-ended.

This open-endedness allows the participants to contribute with much detailed information as they desire; it also allows the interviewer to ask probing questions as a means of following-up. In other words, the interviewees could in principle respond how they like. This can make quite difficult for the interviewer to keep the interviewee on focus while interviewing, and then extract similar themes or codes from the interview transcripts. However, semi-structured interviews reduce individual biases within the study, particularly when the interviewing process involves many participants.

However, this perspective about the risks of qualitative research may lead to a reductionist view that we would like to avoid. Qualitative methods have been described at length in Knowledge Acquisition Processes (KAP) for modelling. Enrico Motta edited a special issue on 25 years of KAP in the Semantic Web area at International Journal of Human-Computer Studies. Elaborating on Gaines, Gruber and Bradshaw’s contributions, he wrote [3, page 132]: “ [...] much of the interesting action concerning knowledge

technologies was actually taking place in the semi-secluded gatherings of this small community and that the real interesting issues were not the formal and abstract Knowledge Representation problems, tackled through “dryerase whiteboard results” (Gruber, this issue), but the ones concerning the effective development of symbiotic intelligent systems (Bradshaw, this issue; Gaines, this issue). These issues could only be tackled effectively through an interdisciplinary approach, grounded as much into empirical investigations and cognitive science principles, as in formal knowledge representation and computational architectures. “

We could not agree more. A genuine non-eclectic interdisciplinarity orientation is key to tackle LYNX problems on building a Knowledge Legal Graph, and to map legal and business requirements.

Hence, we adopted a two-fold strategy: (i) encompassing this empirical approach to properly eliciting modelling requirements across several business and legal fields (as a process); (ii) combining quantitative and qualitative methods in the structured formal line advanced, e.g. by the Unified Modeling Language (UML) perspective (as an outcome). In this sense, completeness, consistency, adequacy, unambiguity, measurability, pertinence, feasibility, comprehensibility, good structuring, modifiability, and traceability will be deemed quality factors to define the goals of the Requirement Engineering process [9, page 35 and ff]. “The requirements emerging from the elicitation and evaluation phases of the RE process must be organized in a coherent structure and specified precisely to form the requirements document “(ibid. 174).

Qualitative research can specify and introduce useful nuances to the summary of preliminary survey results. The interview and focus group techniques based on further elaboration of the previous questionnaire led to interesting results, allowing end-users to refine some of the answers already obtained. First, revealing some internal organizational processes and strategies of government agencies, small / large companies, and law-firms which had not been detected by the survey. Second, providing illuminating expressions and language that summarize the end-user’s conceptual perspective, concerns, and needs on compliance and regulatory problems.

During this phase, 5 interviews and 1 focus group were conducted by researchers within the project. Detailed results for this phase is published in [4] Section 3. Topic classification according to the field notes taken by researchers are:

- **Topic 1:** How legal advisors are searching for relevant information.
- **Topic 2:** How legal advisors prepare relevant information for their lawyers. Identification of the most challenging task of the process.
- **Topic 3:** Accuracy of the information provided.
- **Topic 4:** Information provided to the lawyer.
- **Topic 5:** The need of creating a subsidiary in another Member State.
- **Topic 6:** Suggestions provided by the participants related to the LYNX platform functionalities.

4. Conclusions

This Section briefs the results obtained during the KAP phase within the LYNX project for the development of its platform. For a more detailed information regarding this process, the interested reader could read LYNX Deliverable “D1.1 Functional

Requirements Analysis Report” in [4]. Table 2 plots the functional requirements, as extracted from the surveys:

Table 2. Business requirements extracted from the surveys.

Business requirements regarding the LYNX Platform	Number of mentions
Provide smart references and links among the retrieved documents and any other potentially relevant documents	18
System should exhibit high performance and be able to cope with a very large number of documents	15
Provide summaries of relevant documents	14
Provide smart search services among relevant digital regulatory compliance documents that produce highly relevant results	12
Monitor law, jurisdictions, regulatory compliance and alert users in case of changes, innovations, modifications	12
Provide topic classification within the documents	10
Provide translations of relevant documents	8
Provide recommendations of documents that may also be potentially relevant	7
Include relevant background information and add explanatory information to legal documents so that laypersons are able to understand them	3
Provide access to (at least) the following content areas: tax law, labor law, required permits or necessary authorizations, and operating licenses	2

Table 3 summarizes the expectations of potential end-users of the LYNX Platform, provided through knowledge acquisition techniques (both, quantitative and qualitative) to achieve KAP task. The expectations have been extracted from gains and pains highlighted by the participants in relation to the specific functionalities that the LYNX Platform should provide. Its final goal is to enrich and facilitate the alignment with pilot user’s requirements provided in LYNX Deliverable “D4.1 Pilots Requirements Analysis Report” in [10].

Table 3. General business requirements.

General requirements related to specific features of the LYNX platform (Expectations)	
BR.1	Platform services should be customized according to the professional profile of the end-user
BR.2	Summarization of digital regulatory compliance documents should be provided according to the professional profile: <ul style="list-style-type: none"> • SME, LE, needs to receive specific recommendations related to the relevant regulatory changes occurred within their respective business activity sector. • Consultancy and legal firms need to receive key information related to changes in regulatory compliance with the aim of empowering reasonable and optimal decisions.

- LE Smart search among relevant regulatory documents would be welcome.
- Identifying judgements that involve significant or radical changes in relation to previous legal framework would be useful.
- Identification of key issues removed by the new legal framework with the aim of providing implications of significant changes regarding regulatory compliance.
- Services to perform semantic analysis and linking of content contained within the documents.

PPAs need to provide interpretable legal information

BR.3	Alerts about changes in digital regulatory compliant documents should be provided.
BR.4	Precise translation of digital regulatory documents should be provided.
BR.5	Updating overview of all the applicable regulatory requirements with a link to their documents to support compliance management needs to be provided
BR.6	100% accuracy when setting relevant documents in particular scenarios: providing an accurate classification of documents is really relevant.
BR.7	High-speed updating process is demanded.

References

- [1] P. Casanovas, M. Palmirani, S. Peroni, T. van Engers, and F. Vitali, "Special Issue on The Semantic Web for the Legal Domain - Guest Editors' Editorial: The Next Step," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2765912, Mar. 2016.
- [2] P. Casanovas, J. González-Conejero, and L. de Koker, "Legal Compliance by Design (LCbD) and through Design (LCtD): Preliminary Survey," in *TERECOM@JURIX*, 2017.
- [3] E. Motta, "Editorial: 25 Years of Knowledge Acquisition," *Int. J. Hum.-Comput. Stud.*, vol. 71, no. 2, pp. 131–134, Feb. 2013.
- [4] Jorge González-Conejero, Emma Teodoro, and Pompeu Casanovas, "Lynx D1.1 Functional Requirements Analysis Report," May 2018.
- [5] W. S. Harvey, "Strategies for conducting elite interviews," *Qualitative Research*, vol. 11, no. 4, pp. 431–441, Aug. 2011.
- [6] D. Turner, "Qualitative Interview Design: A Practical Guide for Novice Investigators," *The Qualitative Report*, vol. 15, no. 3, pp. 754–760, May 2010.
- [7] S. Kvale, *Interviews: an introduction to qualitative research interviewing*. Thousand Oaks, Calif: Sage Publications, 1996.
- [8] H. J. Rubin and I. S. Rubin, *Qualitative Interviewing: The Art of Hearing Data*. SAGE Publications, 2012.
- [9] A. van Lamsweerde, *Requirements Engineering: From System Goals to UML Models to Software Specifications*, 1st ed. Wiley Publishing, 2009.
- [10] Julián Moreno-Schneider and Georg Rehm, "LYNX D4.1 Pilots Requirements Analysis Report," May 2018.