An intelligent authoring model for subsidiary legislation and regulatory instrument drafting within construction and engineering industry

CrossMark

Lewis John McGibbney, Bimal Kumar *

School of Engineering and Built Environment, Glasgow Caledonian University, 70 Cowcaddens Road, Glasgow G4 0BA, Scotland, UK

ARTICLE INFO

ABSTRACT

Of primary importance within the domain of open data and more specifically open legislation, lies the essential central requirement for data to be available in a user-oriented manner; whereby the public and professionals alike can consume, share, reproduce it upon request and utilise it when time demands. Focusing specifically on subsidiary legislation (SL), current drafting workflows fall far short of addressing this vision. Whilst a significant amount of recent research has focused on less technical issues such as the actual definition of open data itself, the case for open data as to its management within a legal context, etc. due to the domain specific nature of producing robust and technically accurate open datasets, little work has been done on techniques for the drafting of legislative resources (in particular SL) as open data. To address this problem our work exercises a use-case driven from the domain of sustainable design and construction. As a validation vehicle, we select Scottish building regulations, which govern 32 local authorities across the country as a typical example of such legislation. Our work focusses on three areas of particular importance, (i) observing ongoing practice within the government organisations responsible for the drafting and publication of the aforementioned texts, (ii) understanding the means and methods utilised by local authorities which have been tasked with ensuring that standards of compliance are met within all cases of design and construction, and (iii) reviewing and understanding how construction stakeholders actually execute their activities with respect to the texts in question. The outcome of our study has resulted in a methodology and subsequent production of an intelligent XML authoring workflow model (DROID-SL) for such documents which displays how legal texts of this nature can be better consumed within our society. We demonstrate that by adopting a user-oriented drafting vision, it is possible to produce high quality, user-oriented, linked open datasets which wholly embrace the fast moving area of open legislation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The formal representation of UK SL including UK Statutory Instruments (UKSI's), Scottish Statutory Instruments (SSI's), Welsh Statutory Instruments (WSI's) and Northern Ireland Statutory Rules (NISR's) currently creates huge barriers for building professionals to properly and accurately engage with legal knowledge embedded within the construction-related documents which fall within these legislatures.[1] This practice is not limited to the United Kingdom legislative system; other prominent examples include United States administrative law [13], which can be further refined to the executive, legislative and judicial branches of the US Federal government [1], where "It is

Congress that grants general and specific powers to various Federal agencies through enabling legislation as well as the general laws for their fair and orderly administration. These executive powers are often quasi-legislative in nature (via rules and regulations applicable to a class of persons or organizations) or quasi-judicial in nature (via orders, adjudications and decisions involving particular persons or organizations)" [19], and the legislature of the European Union; in which "One of the defining features of the community and, to a certain extent the Union, is the scope and level of power given to its institutions." [25], where legislative power is disseminated among the Institutions of the European Union. In the latter example, an SL can be produced by either of the European Commission, Council and Parliament which includes acts and agreements by the legislature of the EU. One must, therefore, consider the unique nature of such legislatures, accompanied by the underlying inherited complexity of their historically developed working ecosystems when attempting a contextual study of their legislative outputs. Only then can one observe the dated production level, drafting workflows, highly dependent on technologies which are ignorant (or unaware) of advances made in the domains of legal reasoning, knowledge and ontology representation

---

* Corresponding author.
E-mail addresses: lewis.mcgibbney@gcu.ac.uk (L.J. McGibbney), b.kumar@gcu.ac.uk (B. Kumar).

[1] Throughout this work, we refer to the term legislature as we intend to denote a parliament, assembly and/or elected body of people's representatives. Usually one would expect such an entity to comprise of a group(s) of person(s) with the ability to debate, edit, amend (or have amended) criminal or civil law within the state to which they act. Finally we refer to the relative output (legislative output) as legislation.

[9,24], deontic notions, normative modalities, rights, factors, values and legal rule mark-up and argument making.

An overarching problem is that these administrative workflows, proliferate society with huge amounts of unstructured data, subsequently creating systematic problems which filter themselves throughout many legal, ethical and social aspects of our modern states. Unfortunately, efforts to mitigate against these problems have resulted in minimal measurable impact to date with the construction industry being no exception. Numerous studies [6,20] within the fields of applied computing and automation in construction have focused on improving compliance checking of building codes and regulations, with excellent research published on the topic of semi/fully-automating design and compliance checking [7,11]. This, however does not detract from the underlying debate that current administrative workflows are

- Inefficient; this is the result of the archaic nature of legislative drafting and the dated authoring procedures prevalent throughout legislative ecosystems.
- Highly inconsistent between parliaments and organisations responsible for dealing with the delegated tasks of producing legislation, and finally
- The primary cause of the fact that we, as humans, communicate with, infer from and use legislative documents which are represented in static forms differently every time we put them into application. Ironically, with this in mind, from a political perspective the parliamentary decisions (which include proposals, amendments, approvals, etc.) concerning ever growing quantities of legislation fail to acknowledge or consider the scale of problem the information overload phenomena has on the compliance process overall.

With such topics of international as well domestic importance e.g. Energy Conservation, Environmental Responsibility and Climate Change, we must address the requirement for a more evenly balanced enforcement model which acknowledges factors relating to compliance assurance instead of the existing narrowly directed, bureaucratically dictated model which is heavily biased towards obtaining (on many occasions) politically driven results. Subsequently there exists a requirement to look towards innovation, towards the use of emerging as well as established technologies from across the domains of Artificial Intelligence (AI), Informatics, the Semantic Web, etc. in an attempt to uncover possible solutions to rebalancing the enforcement of legislation as stated above. We have reached a stage within the research community, where a common consensus exists that the age of artificial intelligence within the domain of construction did not materialise into everything the initial hype promised as argued by [27], "The dream in the early days was that it would eventually become possible to develop a computer system of human level intelligence that would automate many of the tasks of engineers (in other words, completely replace human engineers). Unfortunately, this quickly turned out to be impossible, and the AI boom died before long. However, it also triggered diversified research efforts in the search for 'better paradigm AI techniques' to improve and increase information processing capabilities".

Our research focuses on improving professional and the public access to UK SL with a specific focus on construction and building regulations. Having witnessed first-hand, the current practice relating to the everyday use and application of such documents within several Scottish local authority building control departments, our vision to formulate a clearer and more comprehensive platform to enable efficient and consistent use of the SL for design checking and compliance control was initiated focusing not on systems "that facilitate the browsing and retrieval of regulations by industry practitioners" [4], but instead on the formal representation of the underlying data model itself. Previously our work made a comparative study between two such computational data models e.g. the Crown Legislation Mark-up Language and Akoma Ntoso for the suitable representation of UK Construction Regulations.

The research documented herewith presents our motive and methodology for intelligent authoring and subsequent drafting of such SL. Section 2 provides a summated history of applied computing techniques within the legislative drafting domain introducing relevant concepts covering semantic web research and XML-based drafting workflows. Section 3 refines this narrative focusing specifically on SL drafting within the U.K. context by highlighting key limitations in current drafting workflows. Section 4 then introduces DROID-SL our XML authoring framework which specifically builds on two aspects of our methodology for intelligent processing of SL data. At the programmatic, data modelling level we detail the following (i) the intricate discrepancies presented by the underlying Scottish Technical Standards (STS) [26] data model as presented by, and so commonly encountered within documents in the umbrella SL category (ii) our approach to addressing these particular characteristics during mapping to a target data model through the use of automated text transformation, named entity recognition and term annotation based on industry standard internationally recognised schema and dictionary definitions. Section 5 then concludes this paper.

2. History of applied computing in legislative drafting workflows

Earlier research [22] bestows to us the utopian contrivance that "In the emerging framework of the semantic web (where information can be directly processed by computer according to its meaning), legal documents and in particular legislative documents, are undergoing a fundamental change. Being directed to the Internet, rather than to a print house, such documents need to be identifiable in the web, structured according to document models and enriched with machine processable meta-data." The underlying vision of the semantic web is one with which we are now familiar as researchers from several domains have been working to move from rhetoric to reality in order to achieve this paradigm shift. In the run up to the millennium we witnessed the fusing of several research phenomena [9] e.g. description logics, formal ontology and knowledge representation ontology, which were themselves products of technologies which shared their roots with the anagogic fields of Philosophy, AI and Cognitive Science, and which had evolved and matured from the early days of knowledge representation [24] and expert systems such as DENDRAL [12,14], MYCIN [2,23], etc. which we embraced in the 1970's. This amalgamation of the former (millennium era) technologies mentioned above, directly resulted in what we now refer to as the semantic web. Subsequently, focused research into the development and use of semantic web technologies has resulted in widespread adoption of such technologies within many aspects of our societies, with increasing and positive commitment being shown by government administrations globally. If we observe arguments from a variety of domains like legal informatics [22], ontology representation [9] and computing and automation within construction and engineering [4,20], we see the consensus that, "This is achieved using standards based on XML (the eXtended Markup Language) to express document structures and insert in the documents meta-textual information. XML standards can be supplemented with ontology languages (for specifying conceptual structures), and rule languages (for capturing the logical content of legal rules)." [22].

2.1. Review of relevant studies concerning semantic web and legislation

The first of a dedicated book series providing cutting edge rain commentary which falls on the umbrella topics of artificial intelligence and law[2] openly forecasts that technologies relevant to the

---

[2] Expanded sub-topics of absolute importance are driven overwhelmingly by "recent developments in semantic technologies, Natural Language Processing (NLP), ontologies, Information Retrieval technologies (IR) and the Web 2.0 and 3.0…"