EVALUATING LEGALDOCML AND LEGALRULEML AS A STANDARD FOR SHARING NORMATIVE INFORMATION IN THE AEC/FM DOMAIN

Johannes Dimyadi[1], Guido Governatori[2], and Robert Amor[3]

Abstract: Legal text is typically conveyed in natural language and thus not readily suitable for computer processing. Numerous work-around approaches have been proposed by researchers in the Architecture, Engineering, Construction and Facilities Management (AEC/FM) domain over the last four decades to create computable representations of normative data that can be used to automate some of the processes in the domain. The transition from human-readable text to a structured representation can occur in many possible ways, e.g. through Natural Language Processing (NLP) techniques, manual annotations, or through direct coding. In all cases, however, the human-readable document at the source remains the sole point of reference. Ideally, however, one digital structured representation should also be available and recognised as the single digital point of reference.

Research in the AEC/FM domain has shown that automated compliant building design processes would benefit from a single standardised and manageable digital representation of normative data. Recent efforts in the legal domain have shown promising developments in legal mark-up languages such as LegalDocML and LegalRuleML as emerging open standards for legal knowledge interchange. In this article, we explore the potential of adapting these emerging standards to accommodate specific requirements of the AEC/FM domain.

Keywords: legal knowledge model, normative information, automated compliance audit, legaldocml, legalruleml

1 INTRODUCTION

Normative information used in the Architecture, Engineering, Construction and Facilities Management (AEC/FM) domain is a subset of legal knowledge and includes provisions in the primary legislation such as the building act, and subsidiary legislations such as building regulations and building codes, by-laws, as well as standards. The latter may be normative or informative depending on the context and applications.

The AEC/FM domain is still largely practising the conventional manual paper-based information exchange, particularly in the area of compliance audit. Although there have been some promising approaches reported in recent years on automating compliance audit processes, there has yet to be a domain standard for sharing normative information digitally. A common approach has been to encode normative data as rules and either embed them into the programming code or incorporate them as part of the compliance

---

[1] Postdoctoral researcher, Department of Computer Science, University of Auckland, New Zealand, j.dimyadi@auckland.ac.nz
[2] Senior principal researcher, Data61, CSIRO, Australia, guido.governatori@data61.csiro.au
[3] Professor, Department of Computer Science, University of Auckland, New Zealand, trebor@cs.auckland.ac.nz

audit system. One drawback with this approach is the lack of transparency and the cost of maintaining the rules in response to on-going amendments (Dimyadi et al., 2016).

buildingSMART International, a global consortium that promotes open standards for information exchange in the AEC/FM domain, recently set up a working group to recommend a domain standard for representing and processing normative information for automated processes (buildingSMART, 2016). Two emerging standards from the legal domain, i.e. LegalDocML and LegalRuleML (described in this paper), have been put forward as candidates for consideration by the working group. This paper describes the criteria used for their candidacy and reports on the initial findings.

1.1 Sharing Legal and Normative Information Digitally

A comprehensive list of approaches suggested by researchers in the AEC/FM domain over the last 40 years to share normative information for automated compliance audit has been reported in the literature (Dimyadi & Amor, 2013).

In the legal domain, efforts to share digital legal information date back more than 50 years when computerised services and document retrieval database systems appeared in the US and UK in the 1960s (Bourne & Hahn, 2003). This was followed by larger database systems in the 1970s for storing and managing public information such as population and fiscal data. The emergence of personal computers in the 1980s saw a decentralisation of information technology use by individual users, which was also the dawn of automation in legal and administrative offices. The popularity of knowledge-based expert systems of the 1980s had a particular appeal to early researchers in legal informatics as evidenced by many expert systems developed in this period. A more widespread computerised information management systems emerged in the 1990s to support parliaments, judicial and administrative offices. The advent of internet during this time gave rise to online legal information access. Figure 1 summarises various international efforts that facilitate online access to legal information (Sartor, 2011).
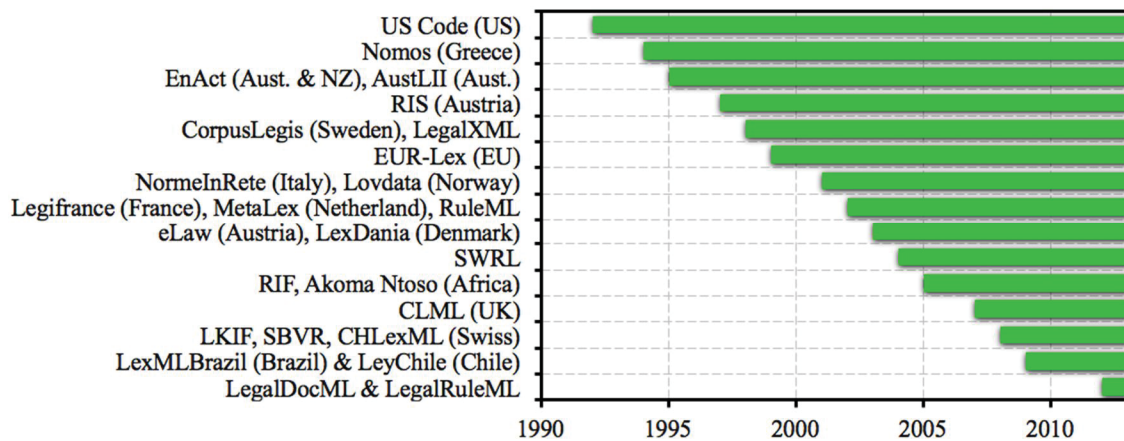

Figure 1: Timeline of computerised legal information models and services

1.2 Approaches to Normative Text Representation

Normative texts in the AEC/FM domain include legislation, regulations and standards, which cannot be processed automatically by machines yet, although there have been some promising NLP (Natural Language Processing) techniques suggested by researchers (Dragoni et al., 2016). Being able to represent natural language text for human and machine-readable is essential for automated processes. As well as to enable computer processing, keeping digital legal documents human-readable helps maintain the user

familiarity with the look and feel of the source paper-based documents, which would promote their adoption in practice.

There have been several rule-based systems suggested by AEC/FM researchers, e.g. SWRL (Semantic Web Rule Language), N3Logic, DRL (Drools Rule Language), etc. These systems focus entirely on the logical content of legal documents without consideration on maintaining a close link with the source documents. This approach lacks transparency and creates a steep learning curve for the end users, which may hinder its uptake by the industry. More importantly, most of them are based on first-order logic (FOL), which can be problematic in representing norms (Herrestad, 1991).

There are two aspects to any document, namely the literal content (text, structure, presentation), and the logical content (or semantics). A digital representation of the literal content is important to human as it maintains familiarity with the paper-based source document. It is equally useful for machines to navigate. The logical content is typically represented as a set of rules designed for machine processing. These rules are also important to the human as they need to be managed and maintained. More importantly, maintaining the same version of both digital and source documents is crucial to ensure the rules in the digital version reflect the latest amendments. None of current rule-based systems supports this kind of association with the source.

1.3   Legal Framework in the AEC/FM domain

There are many legislations, regulations and standards applicable in the AEC/FM domain. They have common characteristics and attributes shared by legal frameworks in many countries around the world. However, there are also fundamental differences in the legal framework and compliance audit processes in different parts of the world. For example, New Zealand has one building legislation and one national building code, whereas US, UK, and Australia have national and regional regulations as well as local bylaws that may have overriding effects. Most countries impose prescriptive regulations where, for example, designs must strictly follow specified methods of construction. However, there is an international move towards performance-based regulations as in New Zealand, where every design must demonstrate that it can achieve a specified level of performance objectives during its service life.

2   REQUIREMENTS AND SELECTION CRITERIA

2.1   Requirements for Automated Compliance Audit Processes

There are generally two sources of information involved in automated compliance audit processes in the AEC/FM domain, namely the building information model (BIM) representing the building design to be audited and the normative information model (NIM), which is used in this paper to describe a collective computable representation of regulations and normative standards. A typical compliance audit process is to check that the specification of certain objects and properties in BIM are in accordance with the requirements, conditions or constraints stipulated by provisions in the NIM. The check can be undertaken progressively throughout the design phase or as a single process at the final design approval or consenting stage.

2.2   Types and Characteristics of Normative Information

The types of normative information applicable for processes in the AEC/FM domain are:

1.   Simple facts, discrete data and constants for use as design parameters

2. Tabulated and graphical lookup data
3. Mathematical expressions (explicit or implied in the text)
4. Explicit rules or conditional statements embedded in the text
5. Deontic operators (obligation, permission, prohibition, and rights) and imperative statements
6. Definitions, commentary and informative (non-normative) text

Some of the rules, deontic operators and imperative statements also have temporal constraints and associated consequences such as penalties or fines.

2.3 Candidacy Selection Criteria

The basic criteria set out to evaluate the suitability of a candidate standard for sharing normative information digitally in the AEC/FM domain are given in Table 1 below.

Table 1: Candidacy Selection Criteria

#	Description of Criteria
C1	The representation must be both human and machine-readable and manageable independently from any system that uses the representation
C2	There must be a close link between the digital model and the paper-based source document to enable automatic version control. This also maintains user familiarity with the structure and literal content of the source document, which is important to promote its adoption in the conventional practice
C3	Availability of practical authoring tools to support the development work
C4	The representation must be based on an open standard technology that promotes interoperability, and supports open standard query languages
C5	The approach must be scalable and have the capability to handle any type of document including normative and informative such as standards, brief, and requirement specifications

3   LEGALDOCML & LEGALRULEML

OASIS (Organisation for the Advancement of Structured Information Standards), an international e-business standards consortium is currently standardising LegalDocML and LegalRuleML (OASIS, 2016). Both of these emerging standards are affiliated with the open standard LegalXML, which started development in 1998 and later managed by OASIS in 2002. The objective of LegalXML is to enable structuring of legal documents, data systems, and information exchange using XML (eXtensible Markup Language) and related technologies.

LegalDocML and LegalRuleML are intended to be the de facto open standards that can operate together to represent the literal and logical content of a legal document for use by both humans and machines. An important feature of the standards is that each rule in LegalRuleML has a unique key that is associated with the unique Id of the source text in the LegalDocML. This provides a version control to ensure the rule reflects the latest amendment of the source provision. Together, they can serve as a standard computable representation of parliamentary, legislative, judicial, and other documents that can be managed independently and allows query and even some reasoning

capabilities (Athan et al., 2015; Lam, Hashmi, & Scofield, 2016; Palmirani, Governatori, & Rotolo, 2011).

In regard to authoring tools, the RAWE and LIME editors developed at the University of Bologna are currently available for use to develop LegalDocML and LegalRuleML documents (Palmirani et al., 2013).

3.1   LegalDocML

LegalDocML is a standardisation of Akoma Ntoso (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontology), which was originally developed by UN/DESA (United Nations Department of Economic and Social Affairs) in 2004 for e-Parliament services in a Pan-African context (Vitali & Zeni, 2007). Akoma Ntoso was partially inspired by CEN (European Committee for Standardisation) Metalex (developed at the University of Amsterdam in 2002) and has been designed to be compatible with it. The entire set of Dutch regulations has been encoded in CEN Metalex. The UK Statute Law Database has also used CEN Metalex. The latest published version of LegalDocML schema is Akoma Ntoso 3.0 (Cervone et al., 2016). LegalDocML is intended to represent only the literal content of any paper-based document, particularly legislations, gazettes, debate records, judicial documents, standards, and other requirements specification.

Below are several features of LegalDocML that are relevant for this evaluation:

1. LegalDocML is exchanged in open standard XML that maintains the structure of the paper-based document very closely, which helps with promoting the user familiarity with the paper-based document. It can also be rendered in HTML for access on any web browser using a predefined transformation style-sheet (e.g. using XSLT).

2. LegalDocML provides a clear separation of metadata from the official content. It uses the open standard FRBR (Functional Requirements for Bibliographic Records) entity-relationship (ER) model to capture the document metadata, which also improves cataloguing and facilitates searches.

3. It has the ability to capture the entire life-cycle of a document and provides an automatic version control and tight coupling with its LegalRuleML counterpart.

Incidentally, LegalDocML was favoured over CLML (Crown Legislation Markup Language) for its ability to adequately express the AEC/FM domain-specific normative information, particularly with respect to UK subsidiary legislations (McGibbney & Kumar, 2013).

3.2   LegalRuleML

LegalRuleML supports different types of statements to model different aspects of norm modelling. An important distinction in legal documents is between constitutive statements and prescriptive statements. Constitutive statements capture definitions of terms to be used in a legal document, while prescriptive statements provide the conditions under which obligations, permissions and prohibitions are in force. Prescriptive statements are modelled by "If <conditions>, then <normative effect>" rules, where the <conditions> are the conditions of applicability of the rule and the <normative effect> is a "Deontic specification". In LegalRuleML parlance, a Deontic Specification is a Boolean statement in the scope of a deontic operator (e.g., obligation, permission, prohibition). An example of deontic specification is "Obligatory fire exit within 20 meters".

Rules are defeasible in the sense that, norms typically describe baseline cases, and then the document can provide many exceptions to the baseline. Thus, LegalRuleML encodes the baseline conditions with one rule, and then each exception is encoded by a more specific rule. The preference relations between exceptions and baseline conditions are modelled by superiority statements. Finally, LegalRuleML is equipped with penalty and reparation statements. Penalty statements are deontic specifications for the various penalties identified by a legal document, and reparations statements link prescriptive statements to penalty statements describing what penalties applies to violations/breaches of specific prescriptive statements.

4 EXAMPLE REPRESENTATION

Provision C4.3 of Clause C4 "Movement to Place of Safety" of the NZBC specifies the following performance criteria that must be demonstrated in any building design:
"The *evacuation time* must allow occupants of a building to move to a *place of safety* in the event of a fire so that occupants are not exposed to any of the following:

- A *fractional effective dose* of carbon monoxide greater than 0.3

- A *fractional effective dose* of thermal effects greater than 0.3

- Conditions where, due to smoke obscuration, visibility is less than 10 m except in rooms of less than 100 m^2 where visibility may fall to 5 m"

The above normative textual provision can be represented literally and structurally in LegalDocML as shown in the following excerpt:

```
<paragraph eId="NZBC_C4.3">
    <heading eId="NZBC_C4.3_heading">Movement to Place of Safety</heading>
    <content eId="NZBC_C4.3_content">
        <p>The <i>evacuation time</i> must allow occupants of a building...</p>
    </content>
</paragraph >
```

The logical content of the above provision can be formalised into three rules, namely nzbcC43DefaultVisibility, nzbcC43VisibilityException, nzbcC43MinEvacuationTime. The first two rules specify that the minimum visibility for a room should be 10 m unless the size of the room is less than 100 m^2. The third rule defines required evacuation time thresholds based on prescribed conditions.

nzbcC43DefaultVisibility: => MinSmokeVisibility($room)=10m
nzbcC43VisibilityException: Size($room)<100m2 => MinSmokeVisibility($room)=5m
nzbcC43MinEvacuationTime: FireEvent,
$MinTimeSmokeVisibility = min{
    TimeTakenFor(SmokeObscuration($room),MinSmokeVisibility($room)):
    for $room in $Rooms},
$MaxTime = min{
    MinTimeSmokeVisibility,
    TimeTakenFor($FractionalEffectiveDoseCarbonMonoxide, 0.3),
    TimeTakenFor($FractionalEffectiveDoseThermalEffect, 0.3)},
$MoveSafetyTime = max{

distance($PlaceOfSafety,$room)/$EvacuationSpeed: for $room in $Rooms}
=>[Obligation] $MoveSafteyTime < $EvacuationTime < $MaxTime

The first two rules are represented in LegalRuleML by two ConstitutiveStatement elements, which provide the specification of parameters to be used in other rules. In this case, the rules give the specification of the term "MinSmokeVisibility" for a room in a building. On the other hand, the third rule is modelled by a "PrescriptiveStatement". The function of this type of statement is to capture obligations, permissions or prohibition; accordingly, the conclusion of the rule inside a PrescriptiveStatement block is in the scope of a deontic operator. The connection among normative textual provisions and their formal representation is modelled by the combination of "LegalSources" and "Association" elements, as shown in the LegalRuleML excerpt below (see www.arcabim.info/lrml/ for the full version):

```
<lrml:LegalSources>
   <lrml:LegalSource key="nzbcC43source" sameAs="NZcode:NZBC_C4.3"/>
</lrml:LegalSources>
<lrml:Associations>
   <lrml:Association>
      <lrml:appliesSource keyref="#nzbcC43source"/>
      <lrml:toTarget keyref="#nzbcC43statements"/>
   </lrml:Association>
 </lrml:Associations>
```

5   DISCUSSIONS AND CONCLUSION

This paper has described the initial work to evaluate the suitability of LegalDocML and LegalRuleML as a candidate standard to express and share the AEC/FM domain-specific type of normative information in automated processes. As summarised in Table 2, the initial findings indicate that they have the attributes to satisfy the criteria.

Table 2: Initial outcome of evaluation for LegalDocML and LegalRuleML

#	Description of Criteria	
C1	They are independent models that can be managed on their own	✔
C2	In combination, they represent the literal and logical content of a document and are tightly coupled for versioning control.	✔
C3	RAWE & LIME editors are available for authoring and development work	✔
C4	They are open standards supporting data exchange in XML	✔
C5	They can represent any type of document	✔

One criterion that has not been assessed in this initial work is the scalability of the approach and the extent to which a representation in LegalDocML and LegalRuleML can be mapped to representations used by existing processes in the domain. This will be investigated as part of the future work.

6 REFERENCES

Athan, T., Governatori, G., Palmirani, M., Paschke, A., & Wyner, A. (2015). LegalRuleML: Design Principles and Foundations. In *The 11th Reasoning Web Summer School*. pp. 1–37. Berlin, Germany.

Bourne, C. P., & Hahn, T. B. (2003). *A History of Online Information Services, 1963-1976*. The MIT Press, England.

buildingSMART. (2016). *buildingSMART International - Regulatory Room*. Retrieved April 9, 2016, from www.buildingsmart.org/standards/standards-organization/rooms/regulatory-room

Cervone, L., Di Lorio, A., Palmirani, M., & Vitali, F. (2016). Akoma Ntoso 3.0 Schema. *The Akoma Ntoso 3.0 - WD17*. Retrieved October 12, 2016, from http://www.akomantoso.org/release-notes/akoma-ntoso-3-0-schema/

Dimyadi, J., & Amor, R. (2013). Automated Building Code Compliance Checking - Where is it at? In *19th International CIB World Building Congress 2013*. pp. 172–185. Brisbane, Australia: Queensland Institute of Technology.

Dimyadi, J., Clifton, C., Spearpoint, M., & Amor, R. (2016). Computerizing Regulatory Knowledge for Building Engineering Design. *Journal of Computing in Civil Engineering*, *C4016001*(August), pp. 1–13. doi:10.1061/(ASCE)CP.1943-5487.0000572

Dragoni, M., Villata, S., Rizzi, W., & Governatori, G. (2016). Combining NLP Approaches for Rule Extraction from Legal Documents. In *Proceedings of the workshop on "Mining and Reasoning with Legal Texts" collocated at the 29th International Conference on Legal Knowledge and Information Systems*. pp. 1–13.

Herrestad, H. (1991). Norms and formalization. In *Proceedings of the third international conference on Artificial intelligence and law - ICAIL '91*. pp. 175–184. Oslo, Norway. doi:10.1145/112646.112667

Lam, H.-P., Hashmi, M., & Scofield, B. (2016). Enabling Reasoning with LegalRuleML. In J. J. Alferes, L. Bertossi, G. Governatori, P. Fodor, & D. Roman (Eds.), *10th International Web Rule Symposium (RuleML 2016)*. pp. 241–257. New York, USA: Springer International Publishing. doi:10.1007/978-3-319-42019-6_16

McGibbney, L. J., & Kumar, B. (2013). A comparative study to determine a suitable representational data model for UK building regulations. *Journal of Information Technology in Construction*, *18*(February), pp. 20–38. Retrieved from http://itcon.org/cgi-bin/works/show&_id=2011elearning/Show?2013_2

OASIS. (2016). OASIS LegalDocumentML (LegalDocML) TC. *OASIS*. Retrieved April 10, 2016, from https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml

Palmirani, M., Cervone, L., Bujor, O., & Chiappetta, M. (2013). *RAWE: A Web Editor for Rule Markup in LegalRuleML*, pp. 1–13.

Palmirani, M., Governatori, G., & Rotolo, A. (2011). LegalRuleML: XML-based rules and norms. *Rule-Based Modeling and Computing on the Semantic Web*, 298–312.

Sartor, G. (2011). Legislative Information and the Web. In G. Sartor, M. Palmirani, E. Francesconi, & M. A. Biasiotti (Eds.), *Legislative XML for the Semantic Web, Law, Governance and Technology*. pp. 11–20. Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-1887-6

Vitali, F., & Zeni, F. (2007). Towards a country-independent data format: the Akoma Ntoso experience. In *Proceedings of the V legislative XML workshop*. 67–86. Florence, Italy.