
Towards Law Versioning: a double category of XML files

I CONTEXT AND OBJECTIVES

Consolidated texts. Legal texts that stay in effect for an extended period of time often tend not to stay static and undergo a process of *modification*. Some of these modifications are carried by a new legal act changing the text of the first one. Viewed with a computer science lens, such acts read as a list of *patches*. — for instance, consider the Treaty of Lisbon¹, whose article 1 begins by the following:

The Treaty on European Union shall be amended in accordance with the provisions of this Article.

PREAMBLE

1) The preamble shall be amended as follows:

(a) the following text shall be inserted as the second recital:

‘DRAWING INSPIRATION from the cultural, religious and humanist inheritance of Europe, from which have developed the universal values of the inviolable and inalienable rights of the human person, freedom, democracy, equality and the rule of law;’

and continues with hundreds of other modifications. As a given text can be modified by several successive instruments, legal publishers — and sometimes also custodian institutions, raising questions on legal intelligibility [8, 15] — often present the text and its successive modifications as a *consolidated text* that is obtained by applying all the patches to the original text. These patches are written in a semi-formal way: for instances, the patches written through an amendment proposed by a French senator² should follow [12, Section III].

This practice presents an interesting theoretical conundrum: the consolidated text — which is widely read and used in all practical matters — has never been adopted; while the ones that did and so are the only ones to have legal standing — the modifying texts — lay forgotten. Furthermore, the process of *consolidating* the text (that is, applying the many patches to the initial text) is mostly done manually³, is time-consuming [4] and error-prone⁴.

The formal language of law. Law is, by custom, highly structured: sections, parts, articles, recitals... and so fits in structured data schemes, as has been remarked several time (see [1]). Under the auspices of the United Nations, the *Akoma Ntoso* standard [9], an XML [3] schema, aims at describing formally parliamentary, legislative, and judicial documents and has been adopted by institutions across the globe⁵. A law described following this format contains all its metadata and present explicitly its structure.

Versioning systems. Versioning systems, such as `git` [14], are widely used in software development and offer a storage model and interface to manipulate different versions of the same document — the different states of a consolidated legal text obviously fit in these models.

From a theoretical point of view, it has been shown [7] that, in order for common operations (such as merging two sets of modifications) on different versions of the same document to be soundly defined,

¹Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007.

²Different institutions have different guidelines.

³The French administration consolidating legal texts has announced to be developing an artificial intelligence system to do so; while the ministry of finances has supposedly developed an experimental tool (*Legistix*).

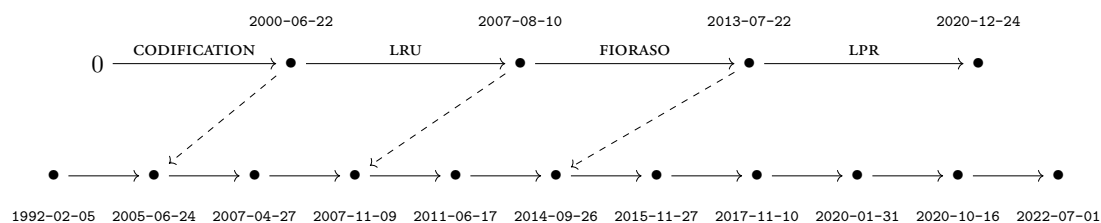
⁴For instance, the *consolidated bylaws* of Université Paris-Est Créteil contain an article 26 *bis* whose vote is not recorded in the minutes of its board of regents, the situation recounted in [10, §2-3] in which the French legislator amended twice (in 2000 and 2016) a law abrogated in 1983 because it was not marked as such in *Légifrance*, or [11] which documents a case where the modifications do not make any sense.

⁵The French Senate publishes all bills and resolutions considered or adopted in this format.

such systems need to consider as primitive elements not the successive states (as `git` does) but the patches themselves (as `darcs` [13] and `pijul` [6] do). Said in more mathematically oriented words, it is necessary to consider the document as a category [5] whose objects are its states and morphisms patches. Nonetheless, for general use, a user of such system is not expected to write patches by themselves, but to let it be computed mechanically (in the form of so-called `diff`-files, that record the lines of a file that were deleted, and the lines that were added) from two subsequent states of a same file, as it is more customary to directly change a file than write a patch to it.

The situation is completely different for legal texts, where the patches are explicitly written and do have precedence from a legal point of view. As the consolidated text is to be computed from the patches, this offers the perfect testbed for a versioning system starting from the patches, which are thus interpreted as *programs* acting on other texts.

Mathematical foundations. We will focus, as a proof of concept, to the bylaws of the Université Paris-Est Créteil — which have been modified fourteen times since 1985, sometimes to follow modifications in texts of higher status — and all the texts that these bylaws depend on (the *Code de l'Éducation* but also other laws and statutory instruments). Together, they form a complex enough but still manageable web of modifications and citations.



The work in [7] is only concerned with completion of the category of files without internal structure and patches. Due to our object of interest, we will need to work in a richer structure: a double category [2] whose objects are XML files, and which have three kinds of morphisms, the *horizontal morphisms* which are patches, the *vertical morphisms* which are citations and the *square morphisms* representing the fact that some citations and patches commute.

The objective of this internship is to define this double category of XML files so as to study whether [7]’s work can be extended to it. It will naturally be divided in many steps:

1. model XML files;
2. define the horizontal arrows;
3. pursue its completion;
4. define thvertical arrows and square morphisms;
5. pursue its completion.

The ideal candidate would have a strong theoretical computer science background, as well as an interest in social science. This work can be followed in many different directions, such as the implementation of such a versioning system.

2 BIBLIOGRAPHY

- [1] Pompeu Casanovas, Giovanni Sartor, Núria Casellas, and Rossella Rubino, eds. *Computable Models of the Law: Languages, Dialogues, Games, Ontologies*. 2008.
- [2] Charles Ehresmann. “Catégories structurées”. In: *Annales scientifiques de l’École Normale Supérieure* **80.4** (1963), pp. 349–426.

- [3] *Extensible Markup Language (XML) 1.1 (Second Edition)*. 2006. URL: <https://www.w3.org/TR/2006/REC-xml111-20060816/>.
- [4] Thierry-Xavier Girardot. “Accéder au droit : importance et défis de la consolidation”. In: *Documentaliste-Sciences de l'Information* (2014).
- [5] Saunders Mac Lane. *Categories for the working mathematician*. 1978, pp. xii+317.
- [6] Pierre-Étienne Meunier and Florent Becker. *Pijul*. 2016-2022.
- [7] Samuel Mimram and Cinzia Di Giusto. “A Categorical Theory of Patches”. In: MFPS. 2013.
- [8] Hervé Moyan. “La consolidation des codes, lois, décrets : positions doctrinales d’éditeurs ou devoir de l’État ?” In: *La Semaine Juridique Edition Générale* (2006).
- [9] Monica Palmirani, Roger Sperberg, Grant Vergottini, Fabio Vitali, and Véronique Parisse. *Akoma Ntoso Version 1.0 Part 1: XML Vocabulary & Part 2: Specifications*. 2018.
- [10] Matthieu Poumarède. “L’évolution de la loi de 1967 relative aux VIC et à l’obligation de garantie à raison des vices de construction : d’une bonne loi à une grande loi”. In: *La Semaine Juridique Notariale et Immobilière* (2017).
- [11] Pierre Sablière. “Mais où est donc passé le 6^e alinéa?” In: *AJDA* (2010).
- [12] Sénat. *La Rédaction des amendements, sous-amendements et motions de procédure*. 2017.
- [13] The Darcs Team. *Darcs, the Darcs Advanced Revision Control System*. URL: <http://darcs.net>.
- [14] The Git team. *Git*. URL: <https://git-scm.com>.
- [15] Jean Waline. “Le grand bazar juridique”. In: *AJDA* (2010).