

Legislation as Linked Open Data: Lessons from MetaLex XML

Alexander Boer

Abstract

Ignorantia legis neminem excusat: ignorance of the law excuses no one. Legislation is arguably the most important resource that a government ought to make available as linked open data, and we are entitled to the highest possible quality of provenance metadata about it. The MetaLex framework sets high standards for publication of legislation and metadata about legislation. This article reflects on the MetaLex standardization process, to identify the most important lessons that can be learned from MetaLex about legislation-as-linked-open-data, aiming to determine what intelligent systems developers should know about legislation, and what they should expect from their government.

1 Introduction

MetaLex is an extensible interchange framework developed between 2002 and 2010 for the XML encoding of legislation. It is designed to be jurisdiction-neutral, language-neutral, and easy to combine with other relevant, open standards. MetaLex, accepted as a European prENorm by the CEN/ISSS in 2006 and 2010 [4], has been discussed in e.g. [3, 5].

In the process of working with stakeholders in the MetaLex CEN/ISSS workshop, from EU member states, a tacit understanding developed about the use cases that were of central importance to users of legislation. This essay reflects on the most important lessons that can be learned from the MetaLex workshop about the nature of legislation-as-linked-open-data.

The workshop members clearly had linked open data in mind: an important objective was to level the playing field for those willing to add value to legislation, made available as open data by governments. National boundaries, and national legal data

silos conforming to national or publisher-specific standards, hold back generic software development for the market of legal professionals. A key feature of MetaLex is therefore its unobtrusiveness in implementation in existing processing environments.

Unobtrusiveness is a result of:

1. the use of an XML metaschema with design patterns, to be extended, rather than a concrete schema to be implemented;
2. the use of a generic naming convention design approach rather than a specific naming convention; and
3. the decision to only require the existence of some extraction mechanism that produces RDF metadata, instead of prescribing some syntactic metadata model.

The best showcase of MetaLex conformance is the Akoma Ntoso standard, developed by UNDESA [9].

Another important objective was to support organizations, often from public administration, who want to manage, and enrich, the legislation of multiple states in the same IT infrastructure. Standardized legislative events play a central role in the MetaLex framework, because we cannot assume that users will intuitively understand the processes behind metadata unless they are explicit. MetaLex standardizes the availability of provenance metadata about sources of law that uniquely identifies resources, and transparently describes the legislative context of production of resources. MetaLex covers any legal-institutional arrangement; president and monarch; civil and common; unicameral, bicameral, and nocameral; decree, act of parliament, and guideline, etc. In an *open* supranational legal information infrastructure, for instance for the growing market of international private law, business-to-government reporting requirements, or tax optimization, a standard like MetaLex is required.

The central theme of this essay is the difference between documents and the legal rules *posited* in them. The corpus of documents and the system of applicable rules do not, and cannot, for practical reasons, march forward in lockstep. To make law transparent, this information has to be conveyed in some way. This subject is of particular interest to those interested in automated legal reasoning. Accurate backing of formalized legal rules by references to legislation, and the dynamic determination of applicability of rules, are essential for maintenance of legal decision support systems in the long run [1]. A solution starts with understanding what is in the corpus, from any vantage point in time.

2 About Legislation and Sources of Law

Let us start by outlining the scope of the MetaLex prenorm. The CEN/ISSS workshop aims for an *Open XML Interchange Format for Legal and Legislative Resources*. The scope statement of the workshop agreement limits the applicability of the XML standard to *sources of law* and references to sources of law [4]. As understood by the workshop, the source of law is a writing that can be, is, was, or presumably will be, used to back an argument concerning the existence of a legal rule in a certain legal system. It is any writing used to communicate the existence of legal rules to a certain group of addressees expected to recognize the legal rules.

The relationship between documents and legal rules is sometimes complex. In law, one often undertakes a formal legal act by producing a written statement in accordance with a certain procedure [1]. In this reading the document is the physical residue of an institutional act: it functions as physical evidence that an act that modified the legal system, for instance through the creation of a legal rule, happened, and it declares the intent of the act, being to create that legal rule. This is why we may refer to an *Act of Parliament*, while we are in actuality referring to the physical result of that act of Parliament. In this sense the documents are a record of changes rather than a snapshot of the current state of the legal rules. Rules can be created by stating them in writing, but modification and repeal of the rule can only be achieved indirectly by way of reference to the legal rule. Moreover, the repeal of a legal rule is not the same thing as its removal from a document; it may be physically present while not being applicable, and it may be applicable while not being physically present in the latest version of the document.

The *source of law* concept covers any lower level regulations, guidelines, or rulings with a legislation-like character that may be used to make claims about the existence of legal rules in a system. Lower level rules too, may be implied by, or recognized in, the legal system. Writings that are clearly precursors, or ingredients, of sources of law are included too, because of their obvious relevance in legal practice. We use the term legislator in the jurisprudential, abstract sense as an entity that produces sources of law, not to be confused with the legislature, or its members.

The concept covers case law in principle, but many MetaLex guidelines make little sense for it. Case law has specific problems: a lack of shared concepts of the provenance of case law, and of the structural elements of the document that carry a posited legal rule that may be cited; the fluidity of the set of judicial decisions that

counts as posited law; and, in some jurisdictions, the practice of referring to abridged summaries of judicial decisions. Judicial decisions play an important part in any legal system, not just in common law systems, but it was less obvious to the workshop what system invariants form a useful basis for additional standardization.

Finally, note that the standardization of references to sources of law makes the MetaLex framework relevant to *any* document that cites sources of law.

3 Problem Context

The MetaLex requirements formulated in the workshop, anticipated linked open data [8]. A number of producers of legislation made their legislation available as XML, or SGML, and published the schema they used, but large institutional users of legislation had little use for this open data, instead relying on the quality of editorial processes of a legal publisher. Legal metadata vocabularies were rarely linked to other metadata vocabularies that might provide context, and a lack of stability of identifiers, so you could reliably point at stuff, was considered a major issue.

Large organizations add metadata and comments to legislation to relate it to their own business logic, and therefore require stable identification of resources over time. Changes may cause considerable, unscheduled work. MetaLex has been received most enthusiastically by professional consumer of sources of law, because it promises them *less* work. Inside governments, one finds producers (legislative drafters, publications offices) and dissatisfied consumers (public service providers) too.

In the legal domain, *intercoder reliability* of XML and metadata encoding procedures matters. Legislation and legislative metadata is rarely all produced by a single government publications office. Usually, legal publishers compete with each other for the most accurate representation of the corpus of legislation. A legal publisher will have its own representation of legislation bundled with its legal information products, and consumers want to merge such products into their own information systems, using agreed upon legislative identifiers as a backbone, instead of linking their business logic directly to commercial products. Ideally, then, different XML and metadata encoders will arrive at the same metadata and XML, with the same identifiers. The room for differences of interpretation of legislative events should therefore be kept to a minimum.

Coverage of all EU jurisdictions in the same information infrastructure remains an increasingly pressing issue, even inside governments. A tax administration in one of the member states must for instance be able to apply the value-added tax regulations

of all other member states. Because foreign users may depend on translations, an infrastructure should be in place that informs about changes, in time, in a way that does not require that one speaks the language, or tacitly understands legislative procedures of other countries.

Importantly, a standard should not require reinterpretation of the entire legal corpus, including its historical record, by experts. MetaLex permits automated translation of existing material, as [8], and unpublished experiences by other government organizations, show.

4 Legislative Metadata

The success of a metadata standard for legislation depends on the extent to which it accommodates maintenance of a continually changing corpus of interrelated documents. A metadata description is about a snapshot of some entity taken in a particular state. A state is a perceived stability of the entity over a particular time interval that does not take account of changes that are outside the domain of interest. For metadata this means in a narrow sense: changes that do not directly affect the data manifestation of the document that the metadata description is about. In a wider sense it means: changes not relevant to the anticipated context of use of a document. The granularity of the snapshot varies across metadata vocabularies, and depends on the targeted community.

To a community that works with legislation daily, the insertion of a new provision is an event to prepare for. For the casual reader, a provision is just one of the constituting parts of a document at the moment of consulting, and what it was before or will later be is of little interest. For the casual reader, and *most* readers, a history and forecast are not required. The casual reader wishes to know what rules are applicable at the moment of consulting. This is a problem that cannot be solved with metadata about the lifecycle of a document, for reasons that will become apparent, but the casual reader is not aware of the subtleties of rule applicability.

The purpose of MetaLex is to provide a basis for legal professionals, who may create metadata about rule applicability, for a specific business context for instance, backed up by an editorial process that generates trust in the metadata. It is not reasonable that the editor of legislation should want to add interpretive, contextualized, and logically complex, metadata to legislation, and MetaLex metadata is intended to be bundled with sources of law.

4.1 Legislative Events

Explicit description of events that change documents plays a vital role in the design principles of the MetaLex framework. It is not unique in choosing this approach: the newer open provenance model vocabulary (OPMV) proposal, is for instance compatible. In [8] both were used in one linked open data server. The drawback of the event description approach is that it results in very verbose models. The MetaLex demonstrator for instance contains over 100 million RDF triples [8], several thousand per document. It is a design decision that should be justified.

In the most popular metadata schemas for documents, single attribute-value pairs are used. Implicitly, these are interpreted as subject-attribute-value triples, where the subject is the document, or a fragment of it. This limitation of the data model makes metadata standards compatible with traditional methods of encoding metadata into HTML and PDF documents. Dublin Core, the vocabulary most commonly used by governments, is an example.

This limited metadata model discourages experts from adding value to the document by annotating it. The problem is that the document to which the metadata is attached is not the proper subject of the metadata. Information about the publication of a document is for instance encoded into a publication-date attribute, with a date value; an author attribute, identified by a string value; and various other attribute-value pairs describing the publication channel. The document itself is treated as the subject this information is about, while these metadata are about the publication of the document, and this intermediate subject is not reified and identified.

Several strong arguments, based on coherence, and use of past and future versions, can be made for the explicit reification and identification of events that happened to legislative documents.

4.2 Coherence

Shared participation in events gives coherence to a corpus of documents. For instance, a previous and current consolidation of legislation, legislation that modified, a modifying authority, legislation that conferred the power to modify to the modifying authority, and a modification date, are naturally joined by a modification event. The modification links together four different documents, and interesting metadata about each of them. Not reifying and identifying events will result in duplication of metadata from different perspectives, unnecessary maintenance, and the loss of relevant references between documents.

Legislative events also offer other important clues about the topical structure of the corpus. In the continental European tradition, legislation that confers the power to legislate for some purpose, is an important clue to the topic of documents.

4.3 Speculation about the Future

The practice of forecasting what legislation will look like in the future, is more important in legal practice than in other fields. Organizations need to prepare for compliance. Public administration must be ready to provide service. Taxpayers may ponder evasive manouvers. Legislation is circulated in expert circles well before it is enacted, published, or even proposed. Legal expert opinions are written, and read by many, that refer to legislation before it is enacted. These opinions, and the references they make, will often remain relevant as knowledge sources if the legislation is actually published and enacted, and become irrelevant otherwise. A modification will only have effect if it is enacted, but the results of the modification have already been applied and widely disseminated by publishers, even though the modification is still subject to change, and therefore speculative. Before events happen, their metadata description is a speculation.

Government may intend to modify a number of acts, and may have published the modifying acts, without having decided on the date of enactment. Experts will speculate, and write about, a momentous legislative event, happening at some time x in the future, that must be prepared for. In attribute-value-oriented document metadata, it is impossible to express that a series of events on documents will take place on the *same*, yet unknown future date. This leaves two options: either the predicted time connection between these speculative documents is lost, or a fictional date is filled in that may create future maintenance problems, if the fiction no longer holds.

4.4 History

Previous versions that are still applicable for some purpose, or frequently referred to, are as important in law as speculative future versions. A tax administration will for instance routinely work at any point in time with at least three different perspectives: the running tax year, the previous tax year, which is being processed, and the next tax year, which is being prepared. Tax deductibility of an investment may be judged by the rules applied at the time the investment was made, which remain applicable in this sense, to validate a judgment made long ago, for decades after they formally lost

force.

4.5 Semantic Interoperability

The structural coherence of the corpus, and the use among experts of past and future versions, differentiate legislation from other forms of publication, and make event descriptions especially important in this field.

In general, one wishes to establish semantic interoperability between different metadata vocabularies. It is only natural to exploit the fact that some types of entities, like documents, people, organizations, places, and dates, are so frequently encountered that they do not fall clearly into the domain of any particular metadata vocabulary. Events always have a central function in relating these entities. There is however no obvious, standard naming mechanism for individual events: the only way to name an event is to refer to its named participants – documents, people, organizations, places, dates, etc – but these should not take the place of the event.

We exploit the fact that one of the acting participants in the event, may be assumed to record events of that type that it participated in. Legislative documents may for instance be identified by the legislator, a type designation, a date of publication, and an ordinal number on that date. Many publications can be identified in similar ways. The lifecycle of legislation is determined by explicitly articulated legislative procedures with named participants. These are jurisdiction-specific, but they are there to be used.

The MetaLex framework supports jurisdiction-specific extensions in a jurisdiction-neutral manner, based on a limited set of prototypical seminal events in the legislative lifecycle, that can be subclassed in an OWL schema. The framework does not support prototypical events that were absent in one of the considered legislative jurisdictions.

5 Bibliographic Identity

For purposes of reliable reference, a determination of bibliographic identity of sources of law is essential. MetaLex standardizes legal bibliographic identity by focusing on the bibliographic events that determine that identity. This is the *naming convention* that MetaLex requires of conforming implementations [2].

A MetaLex conformant implementation uses some naming mechanism that conforms to a number of rules, and distinguishes the source of law as a published work from its expressions; an expression from its manifestations; and the locatable items that exemplify these manifestations, as recommended by [10] (see Fig. 1). A reference to a source of law, either refers to the work, or to a specific expression of that work.

By default, case law refers to expressions, and legislation to works. Modifications are applied to the expression found at the moment of enactment of the modifying act, and not the one found at the moment of publication. The editors of electronic documents refer to, and edit, manifestations under their control, and dereferenceable hyperlinks identify locatable items.

Each bibliographic item exemplifies exactly one manifestation that embodies exactly one expression that realizes exactly one work. A MetaLex XML document is a standard manifestation of a bibliographic expression of a source of law. The work, the expression, and manifestation have a uniform resource identifier reference, conforming to a naming mechanism, and may be individually addressed as subject of metadata.

MetaLex extends [?] with the lifecycle of sources of law, a succession of consolidated versions resulting from legislative events, and optionally so-called *ex tunc* consolidations. If legislation participates in an event, it participates qua work, expression, or manifestation. Consider for example the enactment of a source of law: the work passively participates, but a manifestation may be modified to add the enactment event description as metadata. Documents, as opposed to legal rules, cannot be undone by legislative events. All consolidations of the source of law, past and future, therefore exist in parallel as bibliographic entities, but a current one, by default, reflects the applicable rules at any point in time.

5.1 Retroactive Document Modification

The *ex tunc* expression is a rare consequence of the possibility of retroactive applicability of a modifying act. Ex tunc expressions are specific to law, with its focus on applicability of rules, and shed light on the applicability of rules. The concept of ex tunc expressions captures *errata corrigere* and retroactive annulment of modifications by a constitutional court. The ex tunc expression is a consolidation whose visibility depends on the position you take on the timeline as an observer. Think of a later event e_1 that causes an earlier event $e_2 \prec e_1$, which modifies one expression into a next one, thus retroactively changing the timeline. A decision d_1 at any time after e_1 happened, should be based on the expression created by e_2 , but any decision d_2 between e_2 and e_1 was based on the original expression (see Fig. 2). Normally speaking one would take the present as a vantage point, which means that e_2 happened, but when one reflects on decisions made between e_2 and e_1 , one should see them from a vantage point between e_2 and e_1 . When one takes the decision is distinct from the matter m on which one decides, which remains unchanged in Fig. 2. Whether the rules posited

in the expression created in e_2 apply to m is a separate matter of rule applicability.

In [7], Governatori and Rotolo explain the logical ramifications of annulment, and more generally an overview of the complexities involved in change of the law. MetaLex recommends fictional legislative events caused by later events to model these cases [3].

In case of *ex tunc* expressions, retroactive legal rules have direct bibliographic consequences. This is a rare occurrence. Applicability of rules is a relationship between the rules and the matter to which they apply. To understand it, we have to distinguish between the vantage point we take in time when consulting legislation, after the fact, and the vantage point in time the participants in the matter had. The difference clearly matters. Retroactive (or delayed) applicability of rules in itself has no bibliographic consequences: a retroactive rule *applies to* matters that happened before the text that carries it exists, but it certainly cannot be *applied* before that text exists.

In portals targeted at casual users, retroactive or delayed rules may be carried into past or future versions of documents. This is a user interface device, like the practice of not showing rules which have been repealed. The disadvantage of these practices is that the document manifestation does not concur with an actual consolidation. It has no correct MetaLex identification, because it is a derivative, not covered by the naming convention.

Bibliographic existence and applicability of a legal rule are commonly confused, even by people experienced in versioning. The problem with applicability-based derivatives, is that these are products of an interpretation of legal rules. Differences in interpretation, even among experts, lead to a low intercoder reliability between XML and metadata encoders.

5.2 Referencing Fragments

References to legislation identify not just documents, but parts: specific provisions, sentences, etc. XML structure and fragment identifiers make this possible for manifestations and items. MetaLex adds fragment identifiers for works and expressions.

References do require interpretation. Movement of text fragments for instance creates ambiguity for the identity of reference targets. Provisions are customarily not renumbered when a provision is added or deleted in many jurisdictions. Even law on paper is a kind of hypertext, and legislators do not want to break links. For example: If the legislator moves the text of article 1 to article 2, and adds new text to article 1, we must assume that it is the legislator's intent to make existing legislation-to-legislation references to article 1 *in the corpus* point to the new sentence. Expert interpretations

of the sentence that is now in article 2 may however have to update the reference to article 2.

In [8] a mixed identification strategy, based on MetaLex identifiers, and hash identifiers inferred from the contained text, is used to determine movement of text. Moving the text may however affect applicability of rules. This may be why the legislator moved text. We may not assume, just because the text is still there, that any rule representation for purposes of automated reasoning will remain unaffected, or that interpretations will still be to the point.

A formalized rule representation for automated reasoning can only remain unaffected by changes to the document if it is a syntactically isomorphic representation of the text fragment, and its applicability is dynamically determined. Because no satisfactory implementation of legal rules in a Semantic Web knowledge representation language exists at the moment [6], an editorial process should be in place for such rules before they can be trusted.

6 Conclusions

The key lesson that can be learned from the MetaLex framework, is that the bibliographic uniqueness of legislation is found in the, sometimes complex, relationship between legal rule applicability and bibliographic existence of sources of law positing legal rules. We cannot address legal rule applicability in a bibliographic metadata standard: it remains an interpretation of a relationship between rules and specific facts in a domain to which one would apply them. The MetaLex framework does disentangle bibliographic existence of law from rule applicability.

The bibliographic simulation of complex legislative change operations, and the modeling of *ex tunc* expressions, justify a model of legislative events. The complexities of deciding which rules are applicable, are mainly of interest to those who want to represent the rules found in legislation for automated reasoning in information systems, or want to link them to business processes. To justify the implementation, and to facilitate impact analysis when legislation changes, traceability to legislation is of central importance. Good version management on legislation, and solid bibliographic identification of legislation, are essential first steps. The MetaLex naming convention requirements, and the metadata schema, are based on this tacit understanding. MetaLex has been combined with Legal Knowledge Interchange Format, or LKIF [5], and depends on OWL DL validation of metadata [2].

The urgency of a shared naming convention mechanism for law in all EU member

states has been increasingly recognized in the EU. The European Case Law Identifier (ECLI) is being implemented in member states, and work on the European Legislation Identifier (ELI) has just started, influenced by MetaLex. Interest in linked open data from the government is pushing redesign of government legal information portals. The current trend is to take from MetaLex, or Akoma Ntoso, a conforming standard [9], the requirements that are relevant from a linked open data perspective: conforming naming conventions and event-based metadata.

Ten years ago the MetaLex framework would only be understood as an XML schema; we are therefore making progress. However, the essence of a domain-specific standard remains a correct understanding of denotative semantics: MetaLex users should first and foremost understand the relationship between legislative procedures, and the individuation of sources of law on the work, expression, manifestation, and item level. Only based on a shared understanding, that can be demonstrated by high intercoder reliability, conformance to a metadata standard and naming convention has the added value that we ascribe to it. Each new user is therefore a test of the standard.

References

- [1] A. Boer. *Legal Theory, Sources of Law, & the Semantic Web*. Frontiers in Artificial Intelligence and Applications 195. IOS Press, Amsterdam, the Netherlands, 2009.
- [2] A. Boer. Metalex naming conventions and the semantic web. In G. Governatori, editor, *Legal Knowledge and Information Systems. Jurix 2009: The Twenty-Second Annual Conference*, volume 205 of *Frontiers of Artificial Intelligence*, pages 31–36, Rotterdam, December 2009. IOS Press.
- [3] A. Boer and T. Engers. A Metalex and Metadata Primer: Concepts, Use, and Implementation. In G. Sartor, M. Palmirani, E. Francesconi, M. A. Biasiotti, and P. Casanovas, editors, *Legislative XML for the Semantic Web*, volume 4 of *Law, Governance and Technology Series*, pages 131–149. Springer Netherlands, 2011. 10.1007/978-94-007-1887-6_8.
- [4] A. Boer, R. Hoekstra, E. de Maat, F. Vitali, M. Palmirani, and B. Ratai. Metalex (Open XML Interchange Format for Legal and Legislative Resources). Technical Report CWA 15710:2010 E, European Committee for Standardization (CEN), Brussels, 2010.
- [5] A. Boer, R. Winkels, and F. Vitali. Metalex XML and the Legal Knowledge

- Interchange Format. In G. Sartor, P. Casanovas, N. Casellas, and R. Rubino, editors, *Computational Models of the Law*, volume LNCS 4884 of *Lecture Notes in Artificial Intelligence*, pages 21–41. Springer, 2008. ISBN: 978-3-540-85568-2.
- [6] T. Gordon, G. Governatori, and A. Rotolo. Rules and norms: Requirements for rule interchange languages in the legal domain. In G. Governatori, J. Hall, and A. Paschke, editors, *Rule Interchange and Applications*, volume 5858 of *Lecture Notes in Computer Science*, pages 282–296. Springer Berlin / Heidelberg, 2009. 10.1007/978-3-642-04985-9_26.
- [7] G. Governatori and A. Rotolo. Changing Legal Systems: Legal Abrogations and Annulments in Defeasible Logic. *Logic Journal of the IGPL*, 18(1):157–194, 2010.
- [8] R. Hoekstra. The metalex document server. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, and E. Blomqvist, editors, *The Semantic Web ISWC 2011*, volume 7032 of *Lecture Notes in Computer Science*, pages 128–143. Springer Berlin / Heidelberg, 2011. 10.1007/978-3-642-25093-4_9.
- [9] G. Sartor, M. Palmirani, E. Francesconi, M. A. Biasiotti, and P. Casanovas. *Legislative XML for the Semantic Web*, volume 4 of *Law, Governance and Technology Series*. Springer Netherlands, 2011. 10.1007/978-94-007-1887-6_8.
- [10] K. G. Saur. Functional requirements for bibliographic records. *UBCIM Publications - IFLA Section on Cataloguing*, 19, 1998.

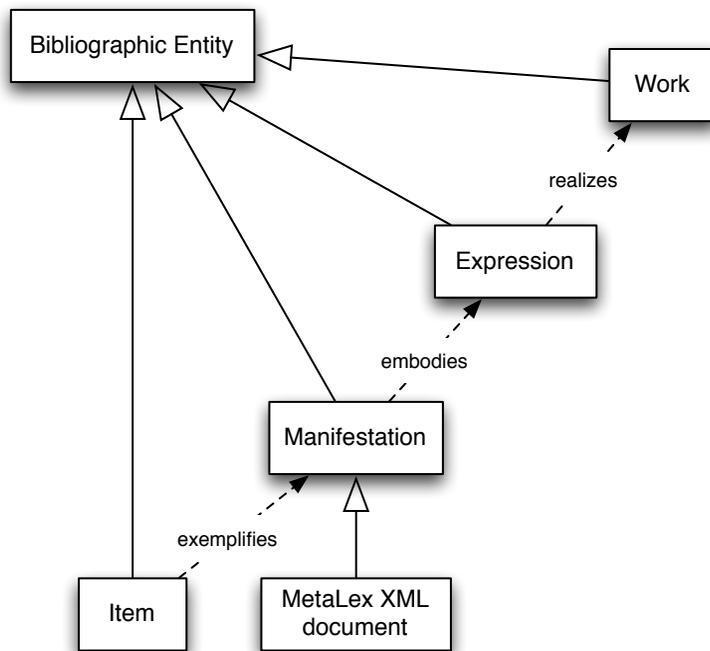


Fig. 1: Taxonomy of bibliographic entities in MetaLex, and relations between them, derived from [10].

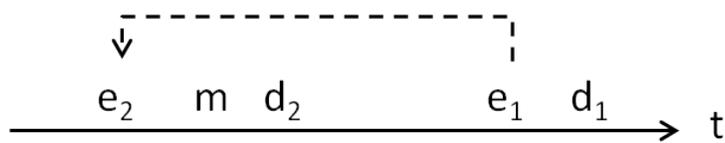


Fig. 2: Ex tunc expressions and vantage points in time t : e_1 causes e_2 retroactively, changing the decision one would make at d_2 to another at d_1 on the same subject event matter m , that happened between e_2 and d_2 .