‘http’ or ‘urn’ URIs for legal resources? How about both?

João Alberto de Oliveira Lima, Monica Palmirani, Fabio Vitali

Legal resources in different national standards have been named using two different identifiers, both backed up by the W3C and IETF: Italian NormeInRete and Brazilian LexML rely on urn-based URIs, while African Akoma Ntoso, Dutch Metalex and many others use http-based URI.
Purpose of this abstract is three-fold: first, we show that http-based URIs have the same useful characteristics that urn-based ones are known to have. Secondly, we show that http-based URIs can, in fact, exceed urn-based ones in terms of flexibility and standard support. Finally, we propose a common framework that can accommodate both approaches (and a few more, including the OpenURLs standard) with conversion paths among them that are easy to implement and to maintain.

1. Naming standards

IETF RFC 3986 (Berners-Lee et al., 2005) is the standard reference for resource identifiers as used on the World Wide Web. RFC 3986 characterizes *identifiers* (i.e., strings containing information useful for distinguishing what is being identified from all other things) for *resources* (i.e., all documents on the web and elsewhere, as well as concrete thing or abstract concept that can be named and is identifiable) that are *uniform* (i.e., independent of their context of use) . The standard contributes to specify that URIs can be *locators* (i.e., they contain information about their current access mechanism such as network address, local name, protocol for retrieval, etc.) or *names* (that are meant to remain unique and persistent and substantially independent of the actual accessibility of the corresponding resource), or *both*, “depending on the persistence and care in the assignment of identifiers by the naming authority, rather than on any quality of the scheme”.
In earlier days of the web, particularly when NormeInRete or LexML were first developed, location was usually associated to http-based URIs (hence called URLs) while persistence in time was associated to urn-based addresses (defined in IETF RFC 2141 (Moats, 2001), and hence called URNs). Contemporary web, on the other hand, has seen a consistent reduction of this distinction, with the abandoning of mere locative properties in many http-based URIs, and the attribution of persistency to some of them. As mentioned in the URI standard and further confirmed by the W3C TAG Finding *URNs, Namespaces and registries* (Thompson & Orchard, 2001), in fact, the persistence of an address, whatever its scheme, is in care of the naming authority, and not an intrinsic property of the scheme in and by itself: http-based URIs can and often are persistent, while nothing except goodwill from the keepers of the urn-based URIs guarantees their persistency.

2. Differences and similarities in using different scheme

To test the flexibility of the http-based and urn-based schemes for addresses in the legislative domain, let us consider the following common scenario:

1. An editor marks up an act in XML (using a standard format such as NormeInRete or Akoma Ntoso). The act contains two references, one to a document that is in fact contained somewhere on the net, while the other has never been converted in electronic format, and thus represents just a virtual reference to a resource that is not accessible on the net. The act is then saved on a public system and made available to the public. For simplicity let us assume that the physical location of this document is unique on the net.

2. A user searches for the document using a search engine, finds it and accesses it by requesting the browser to load it.
3. The document is loaded from the system and displayed on the browser.
4. The user bookmarks the document for a future access.
5. The user activates the references (e.g., by clicking on the corresponding links). The first reference is correctly activated and the corresponding document is shown; the second (pointing to an unavailable document) generates an error detailing the reasons for the failure.

In a urn-based world such as NormeInRete, a reasonable technical solution for this scenario would be the following (please note that neither NormeInRete nor Akoma Ntoso actually explain technical solutions to accompany their standards, so what follows is just our interpretation of a possible solution):

1. The editor determines, according to the NIR URN standard, the urn-based addresses of the two documents, and stores them within the document in a `href` attribute. These addresses could have a form such as `urn:nir:stato:legge:2000-07-09;123`. The document is then saved on a public system and made available to the public.
2. The search engine generates one or more urn addresses corresponding to the document matching the query of the user. These addresses are shown within an HTML page as clickable resources. Since browsers cannot, usually, make use of urn-based addresses, the search engine needs to convert the urn-based addresses into http-based ones. This can be either to the physical locator of the referred document, or, more frequently and wisely, to a resolver service that can, upon requests, retrieve and deliver the actual document wherever it is stored. Although the system displays the urn-based address such as `urn:nir:stato:legge:2000-09-09;234`, the HTML code actually contains an http-based call to the resolver service such as `http://www.resolver.it/cgi-bin/I2L.php?urn:nir:stato:legge:2000-09-09;234`. By clicking on the link, the resolver receives the requests, identifies the corresponding location and redirects the browser (with an HTTP status code such as 301) to the corresponding http-based location.
3. Upon receiving the request, the server where the XML document is stored generates an HTML document to be displayed on the browser (for instance, through an XSLT stylesheet). The document's address displayed in the address bar of the browser **will not be** the urn-based URI, but the current http-based location, e.g. `http://www.camera.it/legge/1234.xml`. Furthermore, since browsers cannot, usually, make use of urn-based addresses, the same XSLT stylesheet needs to convert the urn-based addresses of the references into http-based ones. This, again, is wisely done by adding the http-based location of the resolver service `http://www.resolver.it/cgi-bin/I2L.php?urn:nir:stato:legge:2000-07-09;123`. The reader, when mousing over the anchor of the reference, will be able to see the full http-based reference to the resolver in the status bar at the bottom of the browser.
4. When bookmarking the document, the current location of the document is used, i.e., `http://www.camera.it/legge/1234.xml`, rather than the urn-based URI or even the resolver address, which are completely disappeared from the browser's environment.
5. By clicking on the links of the references, the resolver receives the requests, identifies the corresponding location and redirects the browser (with an HTTP status code such as 301) to the corresponding http-based locations. Again, the

browser will only be able to display in the address bar the http-based location of the document, and not the urn-based URI. Non-existing documents will be identified by the resolver, which will be able to inform the user that no such document exists.

The same scenario in an http-based world such as Akoma Ntoso could be the following:

1. The editor determines, according to the Akoma Ntoso Naming Convention, the http-based addresses of the two documents, and stores them within the document in a `href` attribute. These addresses could have a form such as `/ken/act/2000-07-09/123`. The document is then saved on a public system and made available to the public.
2. The search engine generates one or more http addresses corresponding to the document matching the query of the user. These addresses are shown within an HTML page as clickable resources. Since the http-based addresses are relative, a base URI needs to be established. This could be either the base of the HTML page itself, or placed in the `<base>` tag of the HTML page, or directly inserted in the links to the documents. By clicking on the link, the browser actually activates a link such as `http://www.base.ken/ken/act/2000-09-09/234`. The resolver receives the requests, identifies the corresponding location, and loads the document in order to deliver it to the browser.
3. Upon receiving the request, the server where the XML document is stored generates an HTML document to be displayed on the browser (for instance, through an XSLT stylesheet). The document's address displayed in the address bar of the browser **will be** the http-based URI `http://www.base.ken/ken/act/2000-09-09/234`, and not the current location. Furthermore, the XSLT stylesheet does **not** need to convert the http-based addresses of the references into http-based locations. The reader, when mousing over the anchor of the reference, will be able to see exactly the http-based reference to the documents in the status bar at the bottom of the browser.
4. When bookmarking the document, the exact http-based URI of the document is used, i.e., `http://www.base.ken/ken/act/2000-09-09/234`.
5. By clicking on the links of the references, the resolver receives the requests, identifies the corresponding location, and loads the document in order to deliver it to the browser.. Again, the document's address displayed in the address bar of the browser the http-based URI `http://www.base.ken/ken/act/2000-07-09/123`, and not the current location. Non-existing documents will be identified by the resolver, which will be able to inform the user that no such document exists.

It should be clear now that the http-based solution has no loss of functionality with respect to the urn-based one (i.e., everything that can be guaranteed by the urn-based URI can be also guaranteed by the http-based one), and actually provides a few additional advantages to the urn-based approach:

- The browser displays the http-based URIs and not the physical locator.
- Bookmarks use the http-based URIs and not the physical locator.
- Clickable links use the http-based URIs and not the physical locator.

In short, the original http-based address is maintained in all steps of the procedure, and never substituted with the actual location of the address. The end users and their browsers never even notice that the http-based URI is **not** the physical address. The distinction between physical and logical addresses is completely transparent to the end users, while at the same time maintaining the consistency and persistency

qualities of the urn-based approach (since the http addresses are **not** the physical addresses of the resource).

3. A common framework for http-based and urn-based approaches

In this section we propose a mechanism to solve the problems caused by the two approaches and to allow both to exist and refer to each other by explicitly referring to a common ontological framework that controls and expresses in the same manner all the fragments of both approaches.

3.1. The metadata element set

The metadata element set is the set of relevant pieces of information that are necessary to generate the meaningful fragments of the address.

Our metadata element set is based on the properties of the entities defined in the application of $FRBR_{OO}$ (Doerr & Le Boeuf, 2007) ontology to legal resources (Lima et al., 2007), which is presented briefly in the next paragraph. Here in particular we address the "F46 Individual Work", "F20 Self Contained Expression" and "F21 Complex Work" classes.

While the substance of the "F20 Self Contained Expression" class instance is signs, the substance of "F46 Individual Work" class instance is concepts. Using Palmirani's (2005) terminology, a "normative provision" of a "normative act" is represented by a "F20 Self Contained Expression" instance, and a "norm" is represented by a "F46 Individual Work" instance. In the norm life cycle, one norm could be affected by modifications that generate a new "norm" version represented by one "F46 Individual Work" instance. All "F46 Individual Work" class instances are members of the "F21 Complex Work" related class instance. The "F21 Complex Work" could be used to group norm related component such as annexes.

Table 1 shows the necessary metadata elements to identify one instance of "F21 Complex Work" class. All elements are required if applicable. This table is structured according to some concepts of the "Dublin Core Abstract Model" recommendation (DCMI 2007). In our proposal, we suggest that all typologies (ex.: Norm Type, Norm Content Type etc) and controlled vocabularies (e. g Authority Jurisdiction, Source Authority etc) should be codified using the SKOS (Simple Knowledge Organisation System) recommendation (W3C 2005).

Table 1. Metadata elements for "F21 Complex Work" class.

Element Name	**Type**	**<Indecs> Type**	**Value Surrogate**	**Value String**	**Syntax Encoding**
Authority Jurisdiction	Element	role	non-literal	typed	URI
Source Authority	Element	role	non-literal	typed	URI
Norm Type	Element	type	non-literal	typed	URI
Norm Identifier	Element				
Norm Sign Date	Sub-elem	label	literal	typed	W3CDTF
Norm Identifier	Sub-elem	label	literal	plain	
Norm Component Identifier	Sub-elem	label	literal	plain	

Table 2 shows the additional metadata elements to identify one instance of "F46 Individual Work" or "F20 Self Contained Expression" classes.

Table 2. Metadata elements for "F46 Individual Work" and "F20 Self Contained Expression" classes.

Element Name	Type	<Indecs> Type	Value Surrogate	Value String	Syntax Encoding
Norm Version Date	Element	label	literal	typed	W3CDTF
Norm View Date	Element	label	literal	typed	W3CDTF
Norm Content Form	Element				
Norm Content Type	Sub-elem	label	non-literal	typed	URI
Linguistic Object Language	Sub-elem	role	literal	typed	URI

3.2. A common addressing mechanism

Our proposal specifies five forms of URIs using the specified metadata elements. The Root Form (RF) exposes not only the metadata element value, but also the metadata element name, using a syntax that inspired to the OpenURL standard (ANSI NISO 2004). There are two Base Forms (BF): the BF1 use only the 'http' URI schema, and, the BF2, use only the 'urn' URI schema. Besides these forms, we also define Derived Forms (DF) which uses the Base Forms as parameters of resolutions services.

In the following, assume that *resolver* correspond to the physical address of the resolver service, such as `http://www.base.br` for http-based addresses and `http://www.resolver.br/cgi-bin/I2L.php` for urn-based ones.

Table 3. URIs for Brazilian Act nº 9472 signed at 16/07/1997.

Form	scheme	URI
RF	OpenURL	*resolver*?CountryCode=br&AdministrativeUnitName=federal&NormType=act&NormSignedDate=1997-07-16&NormIdentifier=9472
BF1	http	/lex/br/federal/act/1997-07-16;9472
BF2	urn	urn:lex:br:federal:act:1997-07-16;9472
DF1	http	*resolver*/lex/br/federal/act/1997-07-16;9472
DF2	urn	*resolver*?urn:lex:br:federal:act:1997-07-16;9472

4. Conclusions

In this paper we have introduced an ontologically sound approach to the generation of URI-based addresses for legislative resources that can address the needs of modern parliamentary document systems, and give room to both urn-based and http-based addressing approaches.

Conversions between approaches becomes easy and completely automatic, as long as the correct entities are referenced in both approaches.

It is our intention to suggest the CEN Metalex workshop to consider this proposal for standardization and internationalization.

5. References

ANSI/NISO. Z39.88-2004. The OpenURL Framework for Context-Sensitive Services. Online at < http://www.niso.org/standards/standard_detail.cfm?std_id=783 >

DCMI. Dublin Core Abstract Model. Online at <http://dublincore.org/documents/abstract-model/>.

DOERR, M.; LE BOEUF P. International Working Group on FRBR and CIDOC CRM Harmonization. *FRBR object-oriented definition and mapping to $FRBR_{ER}$ (version 0.8.3)*. 2007. Online at <http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBR_oo_V0.8.3.pdf>.

LIMA, J.A.O; PALMIRANI, M.; VITALI, F. A Time-aware Ontology for Legal Resource. (submitted to ISKO 2008)

PALMIRANI, M. Dynamics of Norms over Time: a Model for Legislative Consolidation. In: *I Quaderni*. n. 18. 2005. pp. 42-69.

W3C. SKOS Core Guide. 2005. Online at <http://www.w3.org/TR/swbp-skos-core-guide/>.

BERNERS-LEE, T.; FIELDING, R.; MASINTER, L. Uniform Resource Identifier (URI): Generic Syntax, IETF Request For Comments 3986. 2005. Online at <http://www.ietf.org/rfc/rfc3986.txt>

MOATS R. URN Syntax, IETF Request For Comments 2141. 1997. Online at <http://www.ietf.org/rfc/rfc2141.txt>

THOMPSON, H.S.; ORCHARD, D. URNs, Namespaces and Registries, Editor's Draft, TAG Finding, 2001. Online at <http://www.w3.org/2001/tag/doc/URNsAndRegistries-50.html>