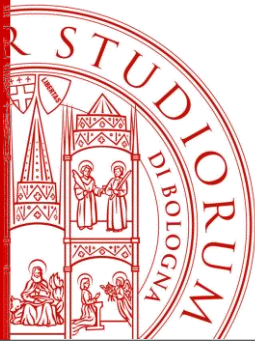


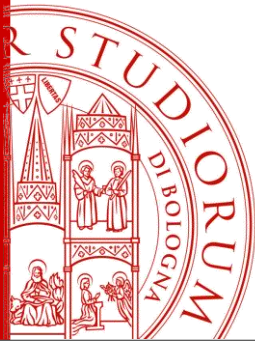
Akoma Ntoso: principi generali

Fabio Vitali & Monica Palmirani
Università di Bologna



Sommario

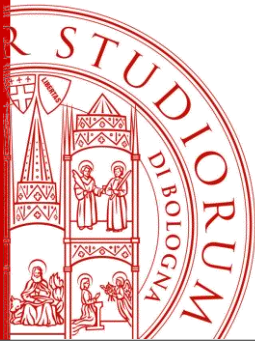
- Si propongono qui le linee guida della progettazione dello standard Akoma Ntoso come risultato delle riflessioni sulle caratteristiche di altri progetti simili (tra cui, importante, quello di Norme In Rete).
- Tra le conclusioni raggiunte:
 - Associazione di quanta più informazione possibile ai contenuti testuali dei documenti
 - Usare non solo XML, ma un XML descrittivo, strutturato, semanticamente utilizzabile
 - Usare URI descrittivi, espliciti, strutturati
 - Favorire la nascita e l'adozione di quanti più software possibili per sostenere al meglio questi standard



Documenti Parlamentari

Utenti e task

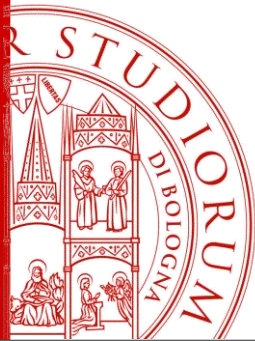
- Drafting legislativo, membri del parlamento
 - Attività di drafting, gestione dei record, traduzione nei vari linguaggi ufficiali e non ufficiali, ecc.
- Back office dei parlamenti: workflow, ciclo di vita
 - Gestione dei documenti nel loro ciclo di vita, memorizzazione, sicurezza, coinvolgimenti di uffici e individui nell'esecuzione dei passi del workflow parlamentare.
- Cittadini: accesso alla legislazione e alle attività parlamentari
 - Pubblicazione multicanale (carta, web, mobile, ecc.) ricerca, classificazione, identificazione
- Tutti: altre attività
 - Consolidamento, confronto tra versioni, sincronizzazione tra linguaggi diversi, ecc.



Documenti Parlamentari

Tool, Tecnologie, o Formati?

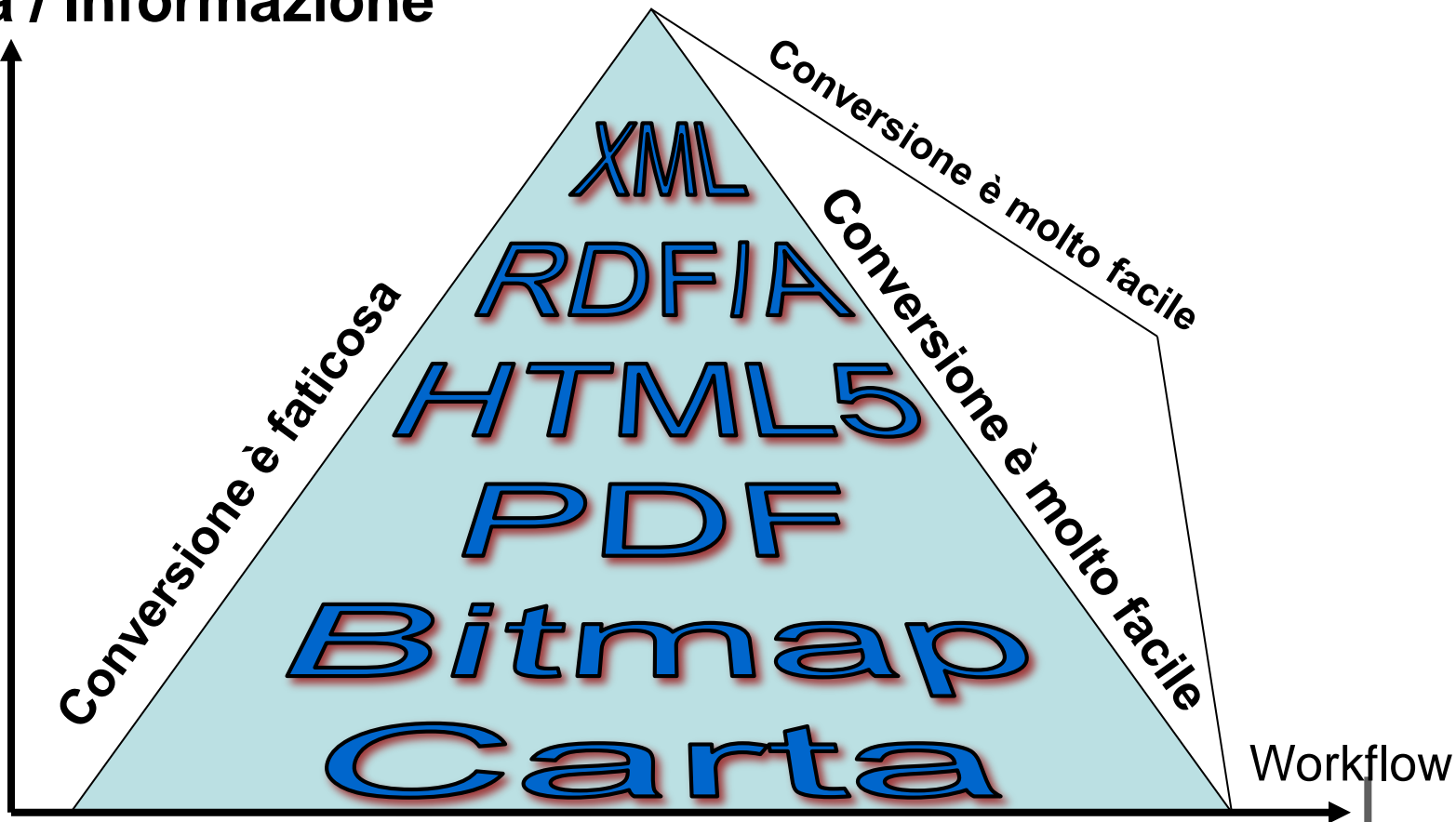
- ***I tool*** si basano su tecnologie concrete, che evolvono nel tempo. Sono i primi a diventare obsoleti, in particolare quando cambiano le tecnologie a supporto.
- ***Le tecnologie*** impongono paradigmi architettureali che a loro volta generano vincoli sulle scelte implementative. Sono più stabili, ma sono influenzati da cambiamenti nell'ambito, negli utenti, nei requisiti.
- ***I formati*** descrivono come le informazioni vengono scambiate tra i tool e come vengono mantenute nel tempo. Più è ricco il dato, più è ampio l'ambito, anche per usi al momento imperscrutabili.

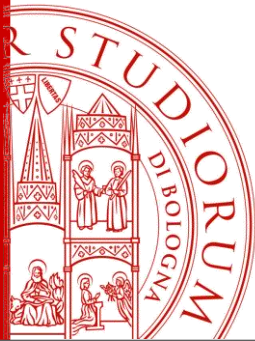


Documenti Parlamentari

Entropia è la perdita di informazioni

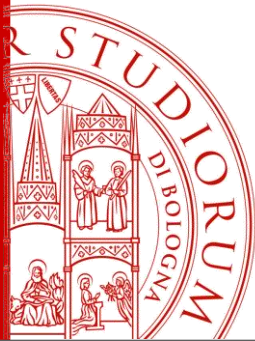
Energia / Informazione





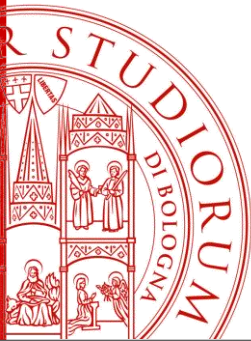
AKOMA NTOSO

- **A**rchitecture for **K**nowledge-**O**riented **M**anagement of **A**frican **N**ormative **T**exts using **O**pen **S**tandards and **O**ntologies.
 - Un approccio strutturato per la progettazione di tipi di documento in XML
 - Un meccanismo sistematico per creare riferimenti ad altri documenti entro e fuori dei confini nazionali basati su URI
 - Un approccio ontologicamente rigoroso per progettare metadati e relazioni tra documenti diversi e versioni diverse dei documenti.
- Iniziato da Dipartimento degli Affari Economici e Sociali delle Nazioni Unite (UNDESA) nel 2006
- Attualmente nelle fasi conclusive di standardizzazione da parte di OASIS, all'interno delle attività del comitato tecnico LegalDocML



AKOMA NTOSO

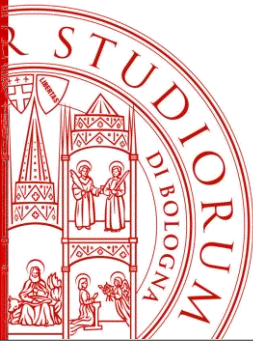
- **A**rchitecture for **K**nowledge-**O**riented **M**anagement of **A**ny **N**ormative **T**exts using **O**pen **S**tandards and **O**ntologies.
 - Un approccio strutturato per la progettazione di tipi di documento in XML
 - Un meccanismo sistematico per creare riferimenti ad altri documenti entro e fuori dei confini nazionali basati su URI
 - Un approccio ontologicamente rigoroso per progettare metadati e relazioni tra documenti diversi e versioni diverse dei documenti.
- Iniziato da Dipartimento degli Affari Economici e Sociali delle Nazioni Unite (UNDESA) nel 2006
- Attualmente nelle fasi conclusive di standardizzazione da parte di OASIS, all'interno delle attività del comitato tecnico LegalDocML



Nome e simbolo

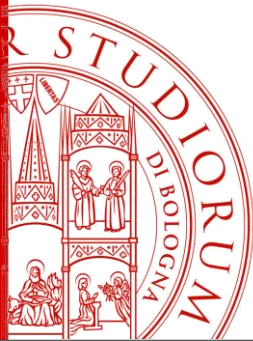


- “Akoma Ntoso” significa *cuori collegati*
- E' un simbolo tradizionale usato dal popolo Akan in Africa Occidentale (Ghana ed altri) per rappresentare comprensione e accordo tra i popoli.
- Similmente, AKOMA NTOSO rappresenta lo standard comune per fornire accesso aperto alla documentazione parlamentare e permettere ai parlamenti di scambiarsi informazioni in maniera più efficiente.



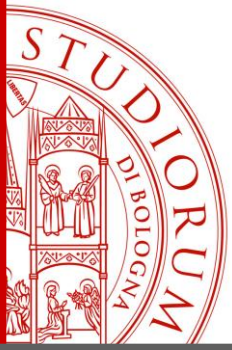
Lo scopo di Akoma Ntoso

- Associare specifiche etichette descrittive (chiamate *markup*) ai documenti legali e legislativi e al loro contenuto.
- Le etichette sono associate all'intero documento, alle sue parti principali, alle strutture secondarie (es. capoversi), fino alle parole.
- Il markup tradizionalmente si divide in:
 - Markup strutturale: identifica le parti del documento e come esse si ripetono, si collegano tra loro e contengono altre strutture.
 - Markup semantico: associa tramite etichette dei frammenti di testo (parole o brevi sequenze di parole) alla loro interpretazione semantica (ad es. il titolo, la data, l'organismo emanante, ecc.)
 - Markup presentationale: associa specifiche istruzioni di visualizzazione (ad es. scelte tipografiche) a specifici frammenti di testo:
- Di grande importanza sono anche i metadati, che per noi sono informazioni associate al documento ma che non fanno parte del contenuto del documento



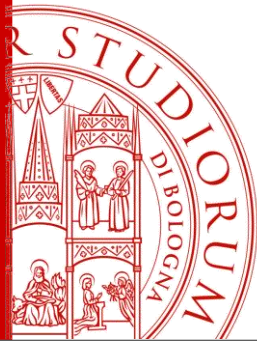
Akoma Ntoso è il canale di entropia minima

- Akoma Ntoso è composto da un vocabolario XML + una URI naming convention
 - La sintassi XML è solo una technicalità. Si potrebbe fare in PDF o in HTML+RDF/A (ma con perdite evidenti di leggibilità)
 - Non è però un XML qualunque, e non potrebbe essere un PDF o un HTML+RDF/A qualunque.
- Akoma Ntoso in realtà è un approccio concettuale per gestire le informazioni associate ai documenti parlamentari:
 - paternità: chi l'ha detto?
 - Differenziazione di contenuto, struttura, semantica, presentazione
 - Precisa concettualizzazione del rapporto tra documento e file
 - Autocontenimento per la preservazione nel lungo periodo.



Design issues per Akoma Ntoso

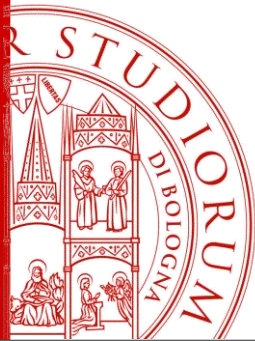
- *Aspetti temporali (di breve e lungo periodo)*
- *Paternità (autori e curatori)*
- *Aspetti dei documenti (struttura vs. presentazione)*
- *Varie idee di documento (FRBR)*
- *Altri aspetti rilevanti*



Aspetti temporali

Per quanto tempo teniamo i documenti?

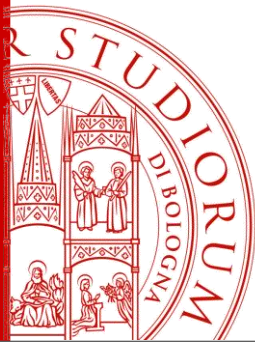
- Aspetti di breve periodo
 - Usarli in molti modi diversi
 - Presentazione multi-canale, workflow documentale, ciclo di vita documentale, consolidamento alla versione attuale, consolidamento point-in-time, ricerche semantiche, classificazione, ecc.
- Aspetti di lungo periodo
 - Usare i documenti tra 5, 10, 50 anni
 - Richiede di non basarsi su alcuna architettura software o sull'esistenza di specifiche risorse, documentazione o manualistica tranne gli stessi documenti...
 - ... Il che a sua volta richiede auto-contenimento, auto-descrizione, set ricco di metadati, markup espressivo, ecc.



Paternalità

chi lo dice? (1/2)

- Dati
 - Il contenuto vero e proprio (testo, struttura, immagini, schemi) come è stato fornito dagli autori e approvato formalmente.
- Metadati
 - Qualunque considerazione, commento o informazione aggiuntiva che possa essere aggiunta al documento o a parti del suo contenuto.
 - I metadati possono essere prodotti sia per intervento umano (da parte di attori diversi dall'autore) sia attraverso processi automatici.
 - I metadati possono essere oggettivi (o, meglio, *indiscussi*) oppure soggettivi: entrambi debbono poter trovare posto nel documento.
- Ontologie (*in breve*)
 - Una rappresentazione formalizzata del modello concettuale che caratterizza i metadati associati ad un documento.

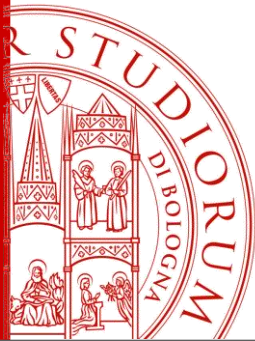


Paternità

chi lo dice? (2/2)

Secondo Akoma Ntoso:

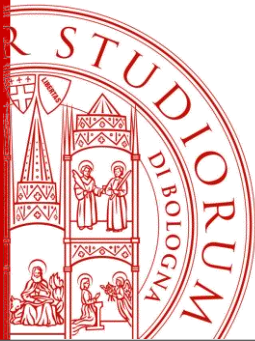
- Il contenuto è sistematicamente separato dai suoi metadati
- I metadati hanno una loro paternità esplicita (anche quelli oggettivi). La paternità del contenuto è implicita (versioni originali) o esplicita (consolidamenti).
- I metadati sono versionati separatamente dal contenuto.
- Collezioni multiple, anche incompatibili, di metadati possono trovare spazio nel documento, associati ad informazioni di paternità
- Markup e metadati, insieme o separatamente, sono aperti a rappresentazioni ed applicazioni semantiche.
- Tecnologie e strumenti del Semantic Web sono immediatamente usabili sui documenti.



Aspetti di un documento

Contenuto, struttura, semantica, presentazione (1/3)

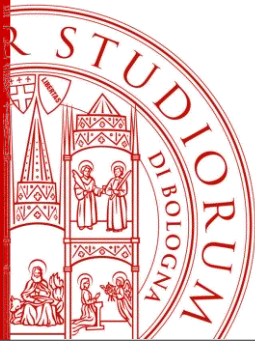
- **Contenuto**
 - Esattamente quello che era scritto nel documento.
- **Struttura**
 - Com'è organizzato il contenuto
- **Semantica**
 - (in breve) il framework concettuale dei dati necessario per comprendere il documento
- **Presentazione**
 - Le scelte tipografiche usate per rappresentare il documento sullo schermo o su carta.



Aspetti di un documento

Contenuto, struttura, semantica, presentazione (2/3)

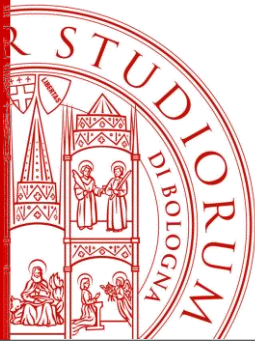
- La struttura aggiunge semantica ai frammenti di contenuto.
 - Le parole “Definizioni iniziali” ricevono significato nel momento in cui sappiamo che sono la rubrica di una sezione nella legge italiana n. 137/2004.
- La struttura collega la presentazione al contenuto
 - Nel momento in cui sappiamo che le parole “Definizioni iniziali” sono la rubrica di una sezione, possiamo applicarvi le scelte tipografiche associate alle rubriche di sezione.
- La struttura può essere usata per testare e validare la correttezza del documento
 - Ad esempio, possiamo dedurre che un documento non è corretto se la sezione manca di una rubrica.



Aspetti di un documento

Contenuto, struttura, semantica, presentazione (3/3)

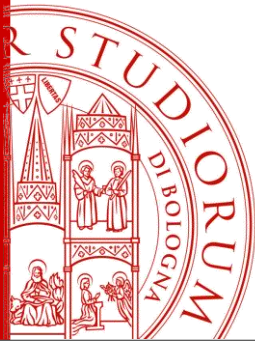
- Possiamo rinunciare a:
 - **Semantica** se dobbiamo solo pubblicare su carta o su schermo
 - **Struttura** se abbiamo bisogno solo di **una** sola scelta presentazionale (ad esempio quella originale)
 - **Contenuto** se abbiamo solo bisogno delle informazioni metadatali (ad esempio se dobbiamo tracciare un documento nel workflow)
- Ma se abbiamo usi molteplici, dobbiamo gestire tutti gli aspetti coinvolti
- Akoma Ntoso ha posto per ognuno di questi.



Varie idee di documento

La gerarchia FRBR

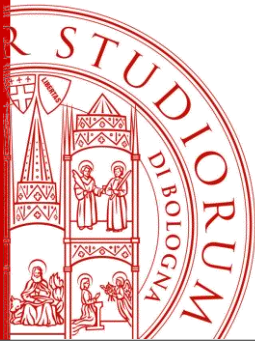
- Concetti diversi
 - Legge italiana 137/2004
 - L'attuale versione consolidata della legge Italiana 137/2004
 - Una rappresentazione in XML della versione consolidata ad oggi della legge Italiana 137/2004
 - Il file “legge137-2004.xml” posto in una specifica cartella su uno specifico computer
- Proprietà diverse
 - Qual è il nome del documento?
 - Chi è l'autore del documento?
 - Qual è la data di creazione del documento?
- La gerarchia IFLA FRBR...
 - **Work**: una creazione intellettuale distinta.
 - **Expression**: la forma specifica in cui si realizza un work
 - **Manifestation**: la rappresentazione di un'expression secondo i requisiti di uno specifico medium
 - **Item**: un esemplare (una *istanza*) di una manifestazione.
- ... fornisce risposte diverse
 - E.g.: un nome diverso per ogni livello
 - E.g.: il legislatore, il curatore, l'editore, il fornitore dei dati.
 - E.g.: la data di promulgazione, la data di consolidamento, la data di generazione, la data della copia



Altri aspetti rilevanti

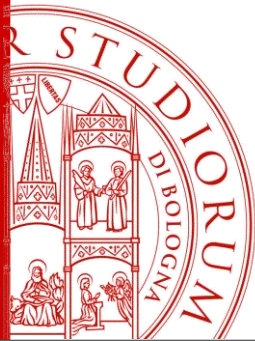
Auto-contenimento

- Il contenuto è testo libero. I metadati sono dati strutturati ed organizzati.
- Strumenti come database e Knowledge Management Systems gestiscono dati, non contenuti.
- Porre i metadati insieme al contenuto garantisce co-manutenzione. Il contenuto e i metadati;
 - Vivono insieme – storage e riferimenti
 - Vengono modificati insieme - allineamento
 - Sopravvivo insieme alle evoluzioni tecnologiche – preservazione
- Akoma Ntoso suggerisce che i tool possano estrarre i dati dai documenti per ragioni di velocità ed efficienza, ma che il documento rimanga la locazione autorevole e permanente dei dati.



Altri aspetti rilevanti gestione del workflow

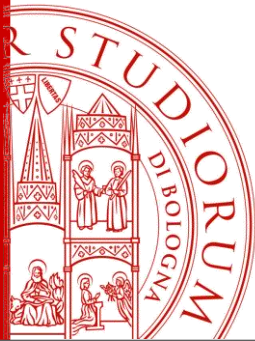
- Il supporto del workflow è un fattore importante di sofisticazione dei metadati
 - Gestione esplicita dell'evoluzione dei documenti
 - Provenance: identificazione delle autorità che emanano cambiamenti (e.g., organi legislativi), fonti di cambiamento (e.g., leggi modificatorie), e specialmente i tempi dei cambiamenti (i tempi degli atti sono una disciplina piuttosto complessa)
 - Identificazione corretta e affidabile di attori e contenuti (incluso l'uso di firme digitali)



Altri aspetti rilevanti

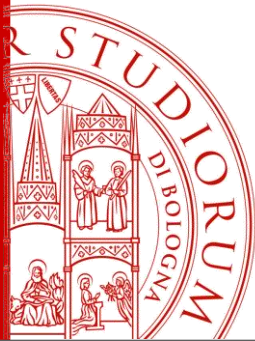
Consolidamento e comparazione

- Possibile solo se struttura, contenuto e presentazione dei documenti sono esplicitamente separati
- Gli approcci tradizionali sono manuali, laboriosi, e richiedono competenza sia legislativa sia tipografica.
- La identificazione esplicita della struttura e l'indipendenza dalla presentazione permette:
 - Il consolidamento come processo semi-automatico basato su riferimenti strutturali espliciti negli emendamenti e atti modificativi
 - La comparazione fianco a fianco come processo totalmente automatico delle differenze tra versione originale e modificata di un testo legislativo.



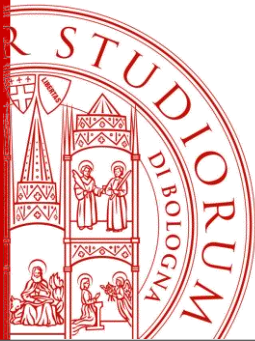
Naming di documenti e frammenti (1)

- Universal Resource Identifiers
 - Usati nel World Wide Web per indicare risorse.
 - I più noti sono gli URL (Universal Resource Locators) usati per navigare sul web
 - <http://www.akomantoso.org/09-examples.html>



Naming di documenti e frammenti (2)

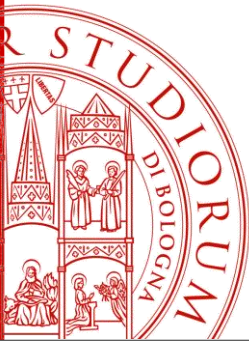
- Coi documenti legislativi, la situazione è più complessa.
- I documenti non sono risorse fisiche, ma entità astratte. Solo i file sono risorse fisiche.
- Tuttavia, i riferimenti legislativi sono raramente (o mai) a file fisici. E' necessario associare URI anche ai documenti, e **debbono** essere diversi dagli URL dei file.
- Il processo di determinazione dell'URL di un item che è la miglior rappresentazione di un riferimento legislativo sottoforma di URI si chiama risoluzione degli URI.
- Un buon schema di naming fornisce una sintassi per gli URI dei documenti e facilita la risoluzione degli URI



Naming di documenti e frammenti (3)

Un buon schema di naming deve essere:

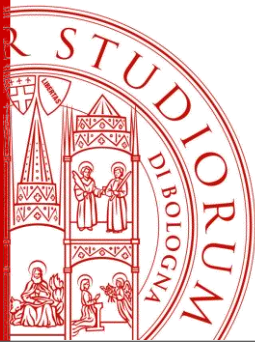
- **Completo:** tutti i documenti rilevanti (in tutti i livelli) debbono essere contemplati
- **Globale:** tutti gli organi legislativi (anche su più nazioni) debbono essere in grado di usarli per identificare tutti i loro documenti.
- **Significativo:** i nomi debbono avere un significato.
 - Permettere assunzioni sul tipo, freschezza e rilevanza di una citazione semplicemente guardando al riferimento.
- **Memorizzabile:** i nomi debbono essere facili da trascrivere, facili da ricordare, facili da correggere se qualcosa è stato scritto in maniera sbagliata.
- **Indovinabile:** dato un riferimento alla legge 137/2004, deve essere possibile dedurre la forma di un riferimento alla legge 76/2006, ecc.



La naming convention di Akoma Ntoso

- Chiara divisione dei livelli FRBR
- Ogni elemento è indirizzabile e individuabile
- Basato su URI con schema http://

Spiegato in dettaglio oggi pomeriggio



Conclusioni

- Uno standard di successo si costruisce su tre fattori fondamentali:
 - Organizzazione precisa e sofisticata di contenuto, struttura, semantica e presentazione
 - Modello completo di metadati (con gestione precisa dei tempi)
 - Modello di naming sofisticato e facile da usare
- NormeInRete, Akoma Ntoso e CEN Metalex ce li hanno
- La regola di base è: non gettare via informazioni che abbiamo e che potrebbero ritornare utili in modi che al momento ancora non conosciamo.