



CNIPA

Ministero della
Giustizia



Il Progetto NIR Standard e Strumenti Software

E. Francesconi

ITTIG-CNR Istituto di Teoria e Tecniche dell'Informazione Giuridica

7 Febbraio 2006

Overview

- Il progetto Norme in Rete
- Gli standard NIR;
- Strumenti software di ausilio all'applicazione degli standard;
- Conclusioni.

The NIR project

- Coordinato dal:

CNIPA

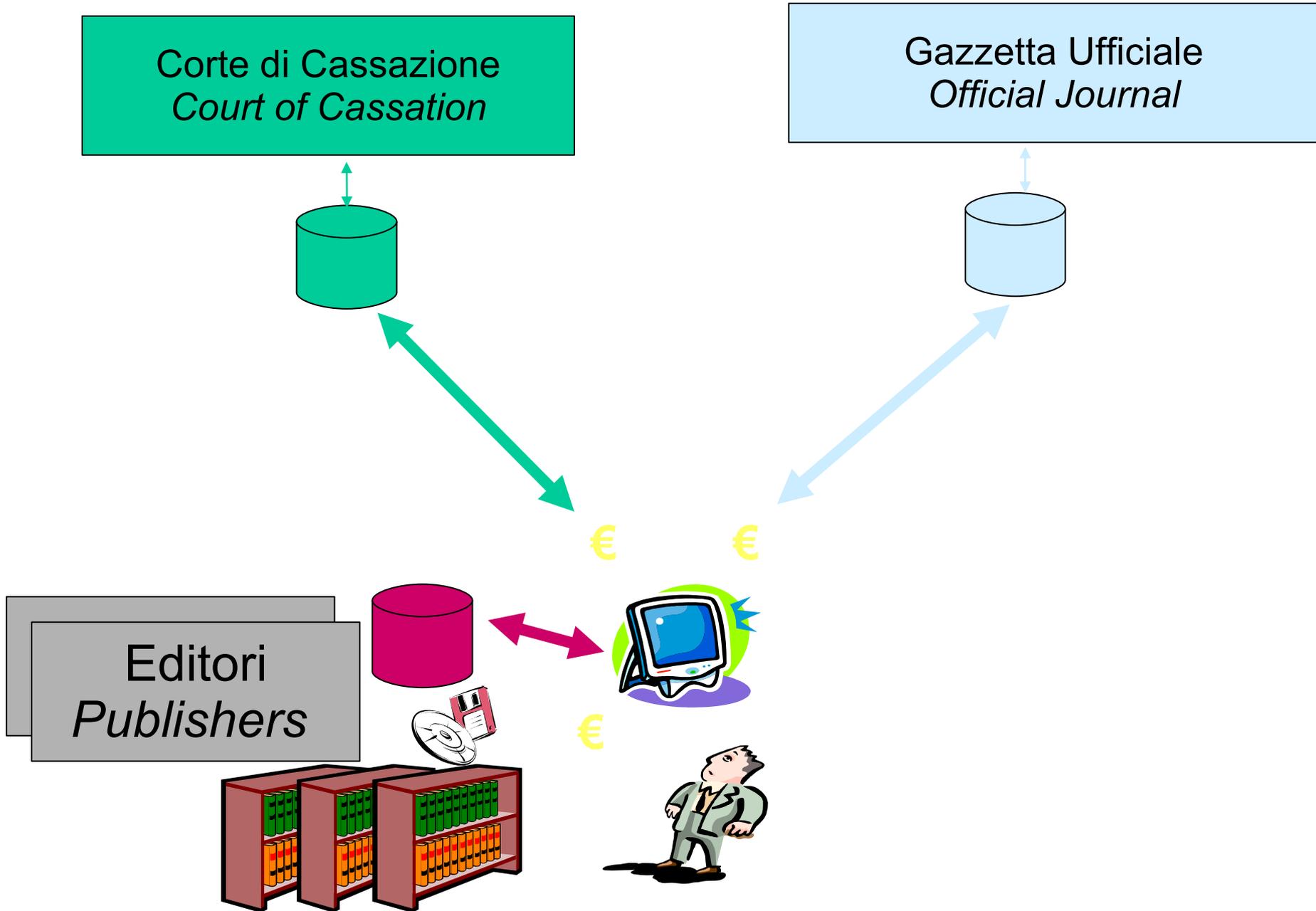
in collaborazione con

**Ministero
della Giustizia**

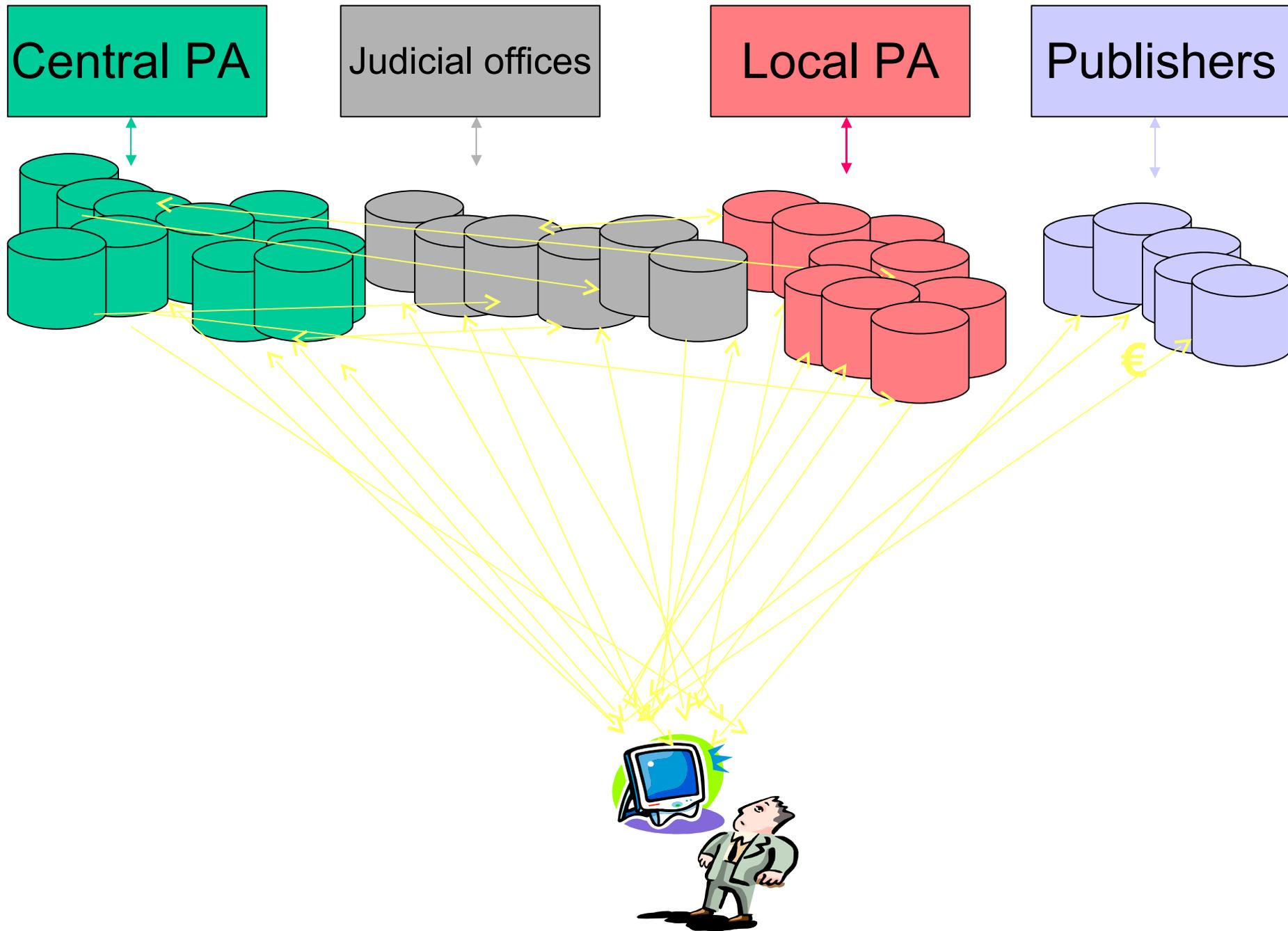
Centro Nazionale
per l'Informatica nella Pubblica
Amministrazione;

- Scopi del progetto:
 - Costruzione di un portale specializzato dotato di funzioni di ricerca per documenti normativi;
 - Guida per l'accesso ai siti Web delle autorità che partecipano al progetto;
- Due linee di sviluppo:
 - 1) Creazione di un punto unico di accesso ai documenti normativi;
 - 2) Meccanismo stabile di referenziazione fra documenti normativi;

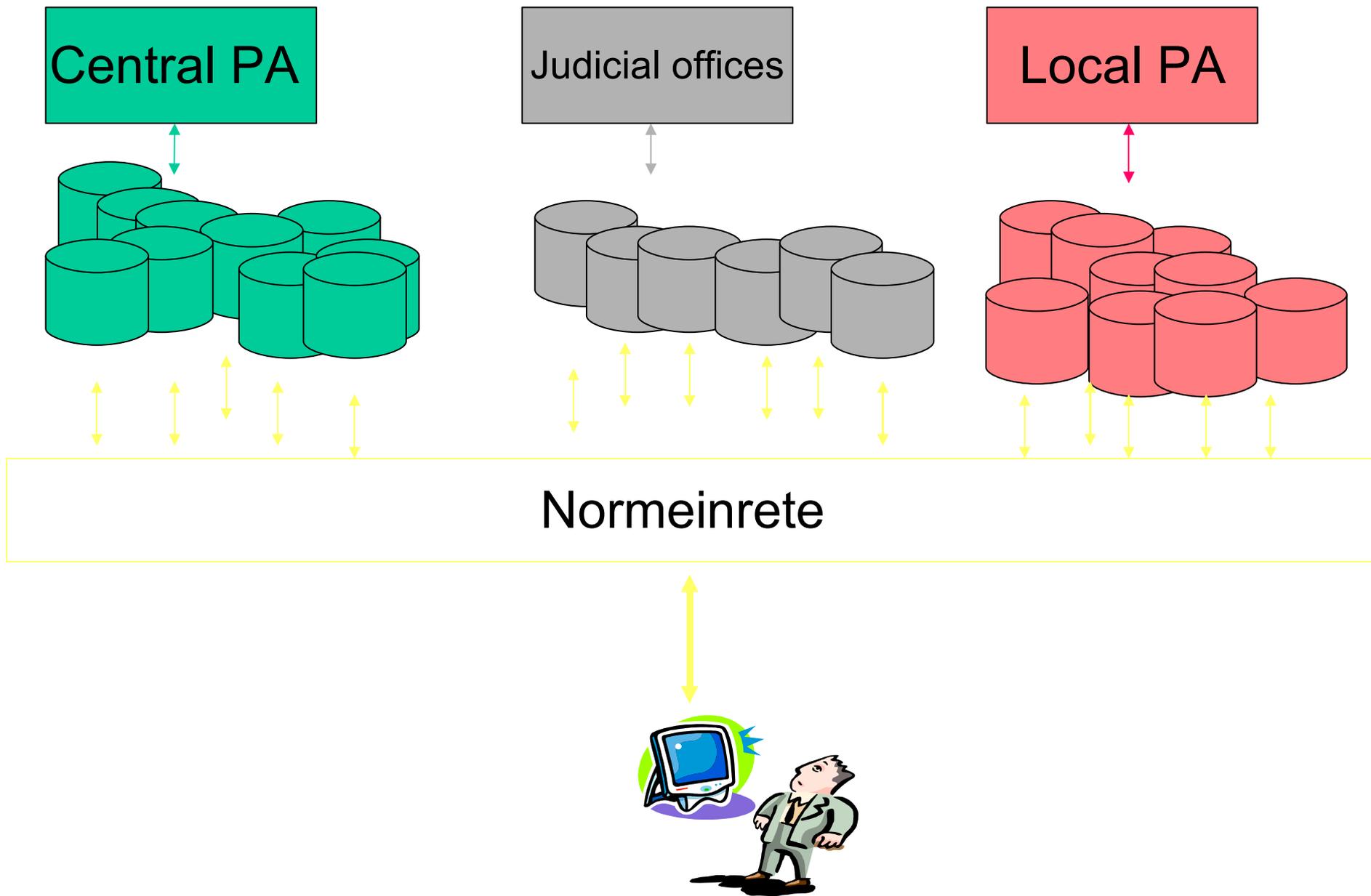
Architetture Centralizzate (80's)



...distribuzione incontrollata (90's)



....Norme In Rete



Gli standard NIR

- Due specifici gruppi nazionali hanno prodotto due standard ufficiali:
 1. Uno standard di identificazione univoca per i documenti normativi basato sulla tecnica dei nomi uniformi (URN):
 - Un identificativo univoco, indipendente dalla localizzazione fisica del documento; esso permette di esprimere i riferimenti fra documenti normativi in maniera stabile, indipendentemente dalla loro localizzazione fisica;
 2. Uno standard XML per la descrizione del contenuto dei documenti normativi definito mediante tre DTD (NIR-DTD) di crescente livello di complessità.

Lo standard URN

- Standard URN-NIR:
 - Schema capace di rappresentare in maniera non ambigua ogni provvedimento normativo, emesso da ogni autorità in ogni data (passata, presente o futura);
 - Un meccanismo di risoluzione da URN a URL;
- Principali proprietà delle URN:
 - Identificano i documenti in maniera univoca e persistente;
 - Sono indipendenti da: disponibilità, modalità d'accesso, localizzazione fisica dei documenti;
 - Non soffrono dei problemi delle URL: perdita di validità, necessità di conoscere la localizzazione fisica dei documenti;
- Le URN sono usate per:
 1. Navigare fra i documenti attraverso i riferimenti (link);
 2. Recupero diretto dei documenti sul Web;
 3. Rappresentazione delle relazioni fra i documenti;

Sintassi NIR-URN

NIR

AUTHORITY

TYPE

DETAILS

ANNEX

- Spazio dei nomi identificato da “nir”;
 - Autorità emanante;
 - Tipo di provvedimento;
 - Estremi del provvedimento: data di emissione, versioni differenti del documento;
 - Annessi;
- **Legge 24 Novembre 1999, No. 468**
`urn:nir:stato:legge:1999-11-24;468`
 - **Decreto del Ministero delle Finanze del 20.12.99**
`urn:nir:ministero.finanze:decreto:1999-12-20;nir`
 - **Regolamento interministeriale dei Ministeri della Comunicazione e della Salute, 9 Settembre 1998**
`urn:nir:ministero.comunicazioni+ministero.salute:regolamento:1998-09-09;nir-1`
 - Per la costruzione della URN sono necessari soltanto i parametri formali di identificazione dell'atto e la conoscenza della sintassi URN;

Hyperlink Automatici

LEGGE 23 ottobre 2003 n. 286

.....

Art. 13 Elettorato attivo

- Hanno diritto di voto per l'elezione del Comitato i cittadini italiani iscritti nell'elenco aggiornato di cui all'articolo 5, comma 1, della legge 27 dicembre 2001, n. 459, che sono residenti da almeno sei mesi nella circoscrizione consolare e che sono elettori ai sensi del testo unico delle leggi per la disciplina dell'elettorato attivo e per la tenuta e la revisione delle liste elettorali, di cui al decreto del Presidente della Repubblica 20 marzo 1967, n. 223, e successive modificazioni.

Hyperlink Automatici

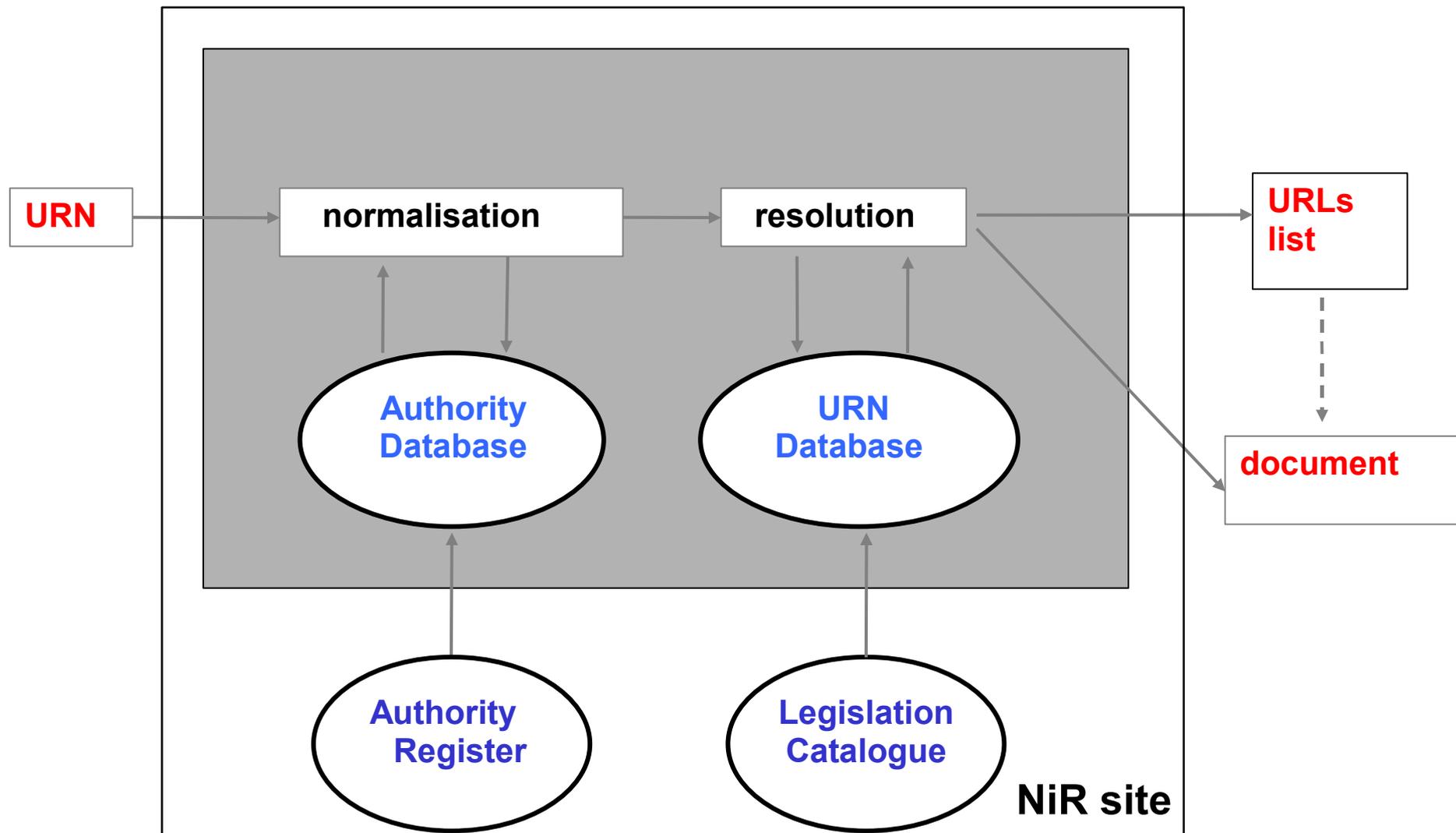
LEGGE 23 ottobre 2003 n. 286

.....

Art. 13 Elettorato attivo

- Hanno diritto di voto per l'elezione del Comitato i cittadini italiani iscritti nell'elenco aggiornato di cui all'
articolo 5, comma 1, della legge 27 dicembre 2001, n. 459, che sono residenti da almeno sei mesi nella circoscrizione consolare e che sono elettori ai sensi del testo unico delle leggi per la disciplina dell'elettorato attivo e per la tenuta e la revisione delle liste elettorali, di cui al
decreto del Presidente della Repubblica 20 marzo 1967, n. 223, e successive modificazioni.

Meccanismo di Risoluzione



Meccanismo di Risoluzione

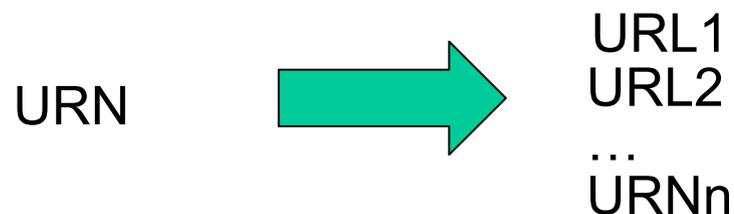
- Normalizzazione:
 - Usa il Registro delle Autorità per la normalizzazione dei nomi uniformi (correzione di possibili errori nella URN di un link)

`urn:nir:ministero.finanze:decreto:2002-09-26`

normalizzazione

`urn:nir:ministero.economia.finanze:decreto:2002-09-26`

- Per consentire il recupero dei documenti, il risolutore lavora anche su match parziali;
- Risoluzione:



Result of NiR Resolution Process



urn

Risultati Ricerca

URN=urn:nir:stato:decreto.legislativo:1997-08-28;281

[Informazioni dal catalogo](#)

Sono state trovate 2 copie del provvedimento rese disponibili da:

- [Parlamento](#)
- [InfoLEGES - Open Community Giuridica](#)

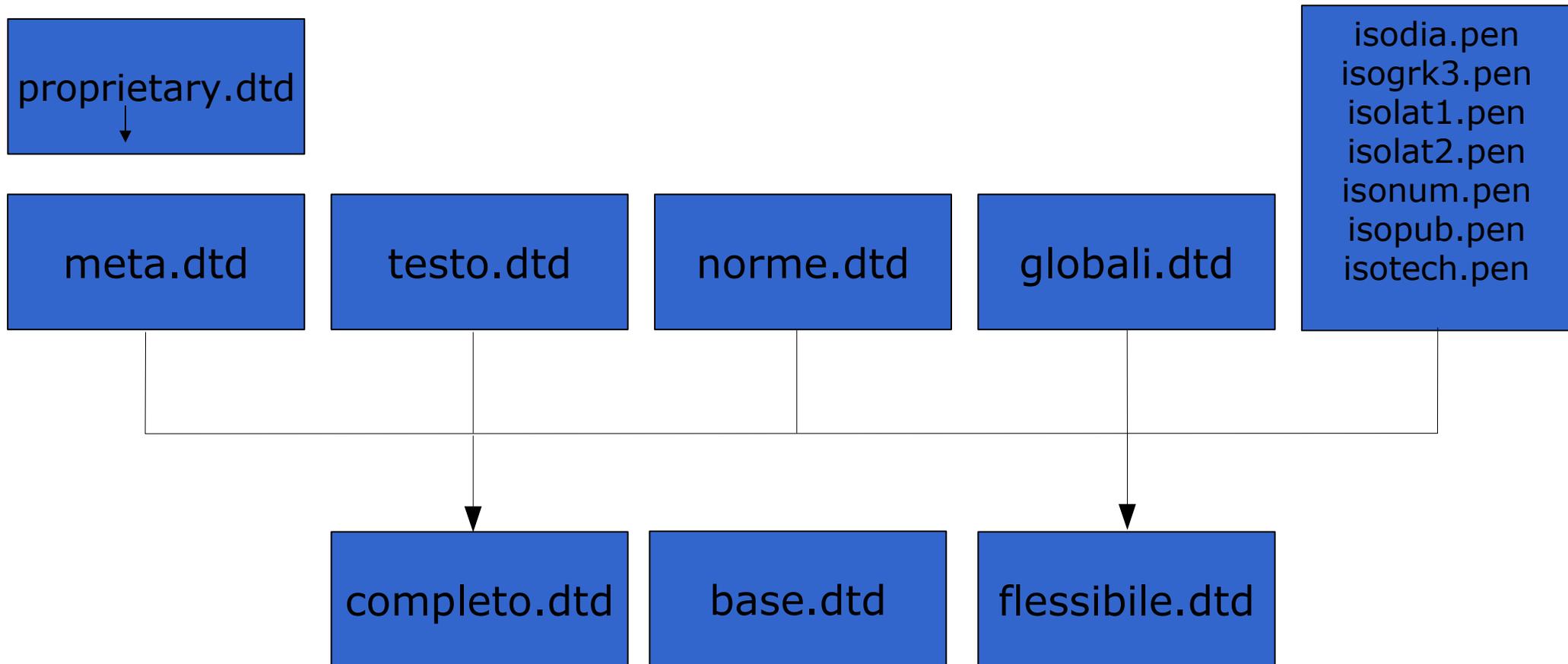
click su un link

2 risorse trovate per un atto

Gli standard NIR-DTD

- Nel dominio NIR i documenti sono descritti da uno standard XML;
- 3 DTD di complessità crescente:
 - “DTD flessibile” (*nirloose.dtd*) (180 elementi):
 - Non ha regole obbligatorie;
 - Usato per documenti pregressi che non seguono le regole di drafting legislativo;
 - “DTD completo” (*nirstrict.dtd*) (180 elementi):
 - Segue le regole di drafting legislativo;
 - Usato per la redazione di nuovi atti;
 - “DTD base” (*nirlight.dtd*) (100 elementi):
 - Rappresenta un sottoinsieme del “DTD completo”;
 - Utilizzato per la formazione all’uso degli standard NIR-DTD.

DTD NIR - Organizzazione



Categorie di elementi XML-NIR

- Elementi strutturali
 - Intestazione, preambolo, sezioni, articoli, paragrafi, ...
- Elementi speciali
 - Riferimenti ad altre leggi, particolari entità (tabelle, liste, ecc.);
- Metadati

Tipi di Metadati

- Metadati Generali
 - Classificazione, Data di pubblicazione;
 - Relazione fra gli atti.

- Metadati analitici (*Modello delle Disposizioni*)
 - “Modifiche”
 - Integrazione, Abrogazione, Sostituzione;
 - “Regole”
 - Obbligo, Divieto, Sanzione, Eccezione, ...

e loro Argomenti;

Descrizione dei documenti da parte delle NIR-DTD

- Le NIR-DTD descrivono un testo normativo secondo due profili:
 - **Profilo formale:**
 - Considera un testo di legge composto da partizioni formali;
 - Descrive la struttura fisica di un documento;
 - **Profilo funzionale:**
 - Considera un testo di legge composto da componenti elementari dette *Disposizioni* (frammenti di regolazione);
 - E' descritto dai metadati analitici composti da tipi di disposizioni e loro argomenti;
 - Descrive la struttura logica di un documento.

Frammento di un testo di legge visto secondo i profili formale e funzionale

Articolo 7
(Notificazione)

1. *Il titolare che intenda procedere ad un trattamento di dati personali soggetto al campo di applicazione della presente legge e' tenuto a darne notificazione al Garante*

1. *La notificazione e' effettuata preventivamente ed una sola volta...*

Profilo Formale

Comma

Profilo Funzionale

Tipo di Disposizione:
"Obbligo"

Argomenti:

Destinatario:
"Titolare"

Azione:
"Notificazione"

Controparte:
"Garante"

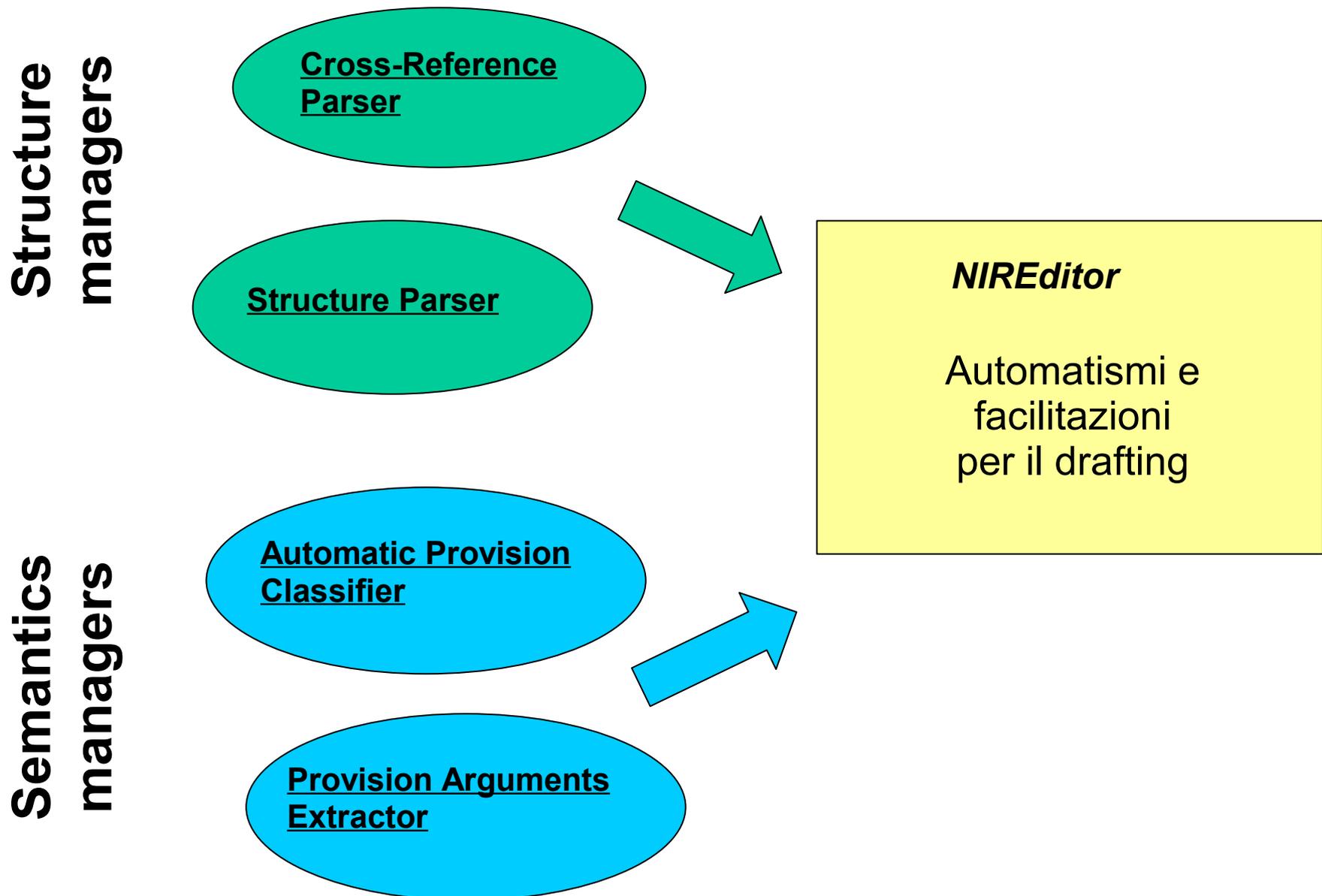
Uso dei Metadati Analitici

- Modifiche
 - Consolidamento;
- Regole:
 - Interrogazioni di un sistema informativo dal punto di vista semantico;
Esempio di query:
“Recupera tutte le disposizioni di tipo “Obbligo” che hanno come destinatario il “Titolare””

Nuova versione delle DTD NIR

- Maggiori novità sui Metadati Analitici;
- Descrivono la dinamica di un testo oltre alla semantica;
- Dinamica:
 - Tempi esterni:
 - Tempi sulle disposizioni (“vigore”);
 - Tempi sulle norme relative (“efficacia”);
 - Tempi interni:
 - Tempi sugli argomenti
(ex: “una imposta deve essere pagata entro [*una data*]”)

Strumenti per l'adozione degli standard NIR



Parser dei Riferimenti

- Progettato per il riconoscimento dei riferimenti e la costruzione delle URN relative in documenti:
 - “plain text”;
 - HTML;
 - XML-NIR (dove i riferimenti non siano stati marcati);
- Il parser è generato usando le tecnologie
 - LEX: estrattore di elementi significativi (token);
 - YACC: generatore di analizzatore sintattico che compone gli elementi estratti sulla base della grammatica dei riferimenti espressa in notazione EBNF.

Parser di Struttura

- Progettato per aiutare la conversione in XML-NIR di documenti preesistenti;
- Due strategie di parsing:
 1. Corpo della legge:
 - Automi a stati finiti non deterministici (NFA)
 - Gli stati sono rappresentati dagli elementi delle DTD-NIR;
 - Le transizioni sono associate a regole formali nella divisione delle partizioni (es: l'inizio di un comma è caratterizzato da)
 2. Formule iniziali e finali:
 - Hidden Markov Model (HMM)
 - I nodi del grafo sono rappresentati dalle parole;
 - Gli stati nascosti associati ad ogni nodo corrispondono agli elementi XML dell'intestazione e della conclusione di un documento NIR.
(Analogie fra sequenze di parole)

Classificatore automatico delle disposizioni

- Progettato per la classificazione automatica di commi nei tipi di “disposizioni” previsti dallo schema dei metadati analitici NIR;
- Sono state sperimentate due metodologie:
 - *Naïve Bayes*
 - *Support Vector Machine*;
- Operano su rappresentazioni del testo in termini di vettore di parole:
 - Sperimentazione sul *plain text*;
 - Testo filtrato mediante tecniche di analisi del linguaggio.

Feature Selection

- Selezione delle *features* significative (parole) per ridurre la complessità del problema;
- La selezione può essere effettuata tramite *Information Gain*: misura la capacità discriminativa di una parola rispetto ad una classe;
- *Information Gain* è calcolato in termini di riduzione di *Entropia*:
 - Fissata una parola e una classe, si verifica il modo in cui si distribuiscono gli esempi che contengono la parola fra quelli positivi (\in classe) e quelli negativi (\notin classe);
 - Il caso ideale è $Entropia(W, D) = 0$ (dispersione nulla, tutti gli esempi che contengono la parola \in classe, e viceversa).

Word Stemming

- Per *Word Stemming* si intende la funzione di normalizzazione delle parole;
- Una stessa termine che si manifesta con generi (per i sostantivi e i verbi) o coniugazioni (per i verbi) distinte si suppone che contribuisca allo stesso modo alla individuazione della categoria del testo;
- Es:
 - “deve” “devono” “dovranno” “dovrà”
si suppone che contribuiscano allo stesso modo nella classificazione di una disposizione di “Obbligo”;
- Gli esperimenti che hanno utilizzato la funzione di Stemming per la lingua italiana hanno dato i risultati migliori.

Esperimenti: *Data Set*

- Sono stati selezionati 582 esempi (*data set*) di disposizioni distribuiti su 11 classi (tipologie di disposizioni);

C₀ Abrogazione

C₁ Definizione

C₂ Delega

C₃ Delegificazione

C₄ Divieto

C₅ Eccezione

C₆ Integrazione

C₇ Obbligo

C₈ Permesso

C₉ Sanzione

C₁₀ Sostituzione

Training e Test del Classificatore

- Training:
582 esempi utilizzati per addestrare il classificatore;
- Test: due esperimenti per calcolare
 1. *Train accuracy*:
 - Risultato del riconoscimento sul training set;
 2. *LOO (Leave-One-Out) accuracy*:
 - Tutti gli esempi sono usati per addestrare il classificatore eccetto uno che usato per il test.
 - Si ripete l'addestramento lasciando un esempio differente fuori dal training set, finché tutti gli esempi non sono stati usati per il test. (Test della capacità di generalizzazione del classificatore)

Risultati Naïve Bayes

- Utilizzo di tutto il testo contenuto nelle disposizioni;

Train accuracy	Loo accuracy	Prime n parole con max infogain
90.7%	86.9%	100
89.3%	86.9%	50

- Eliminazione del testo contenuto nelle virgolette

Train accuracy	Loo accuracy	Prime n parole con max infogain
95.5%	88.6%	500
94.3%	88.1%	250

SVM

- Utilizzo di tutto il testo contenuto nelle disposizioni;

Train accuracy	Loo accuracy	Prime n parole con max infogain
100%	91.2%	1000
100%	91.9%	500

- Eliminazione del testo contenuto nelle virgolette

Train accuracy	Loo accuracy	Prime n parole con max infogain
99.8%	92.1%	tutte
99.8%	92.1%	1000

Chunking

- Rappresentazione del testo mediante strutture linguistiche a più alto livello di astrazione;
- Compressione della rappresentazione del testo mediante:
 - Rappresentazione simbolica di sottostrutture grammaticali;
 - Rappresentazione simbolica di sottosequenze;
- Collaborazione ITTIG-ILC.

SVM con chunker

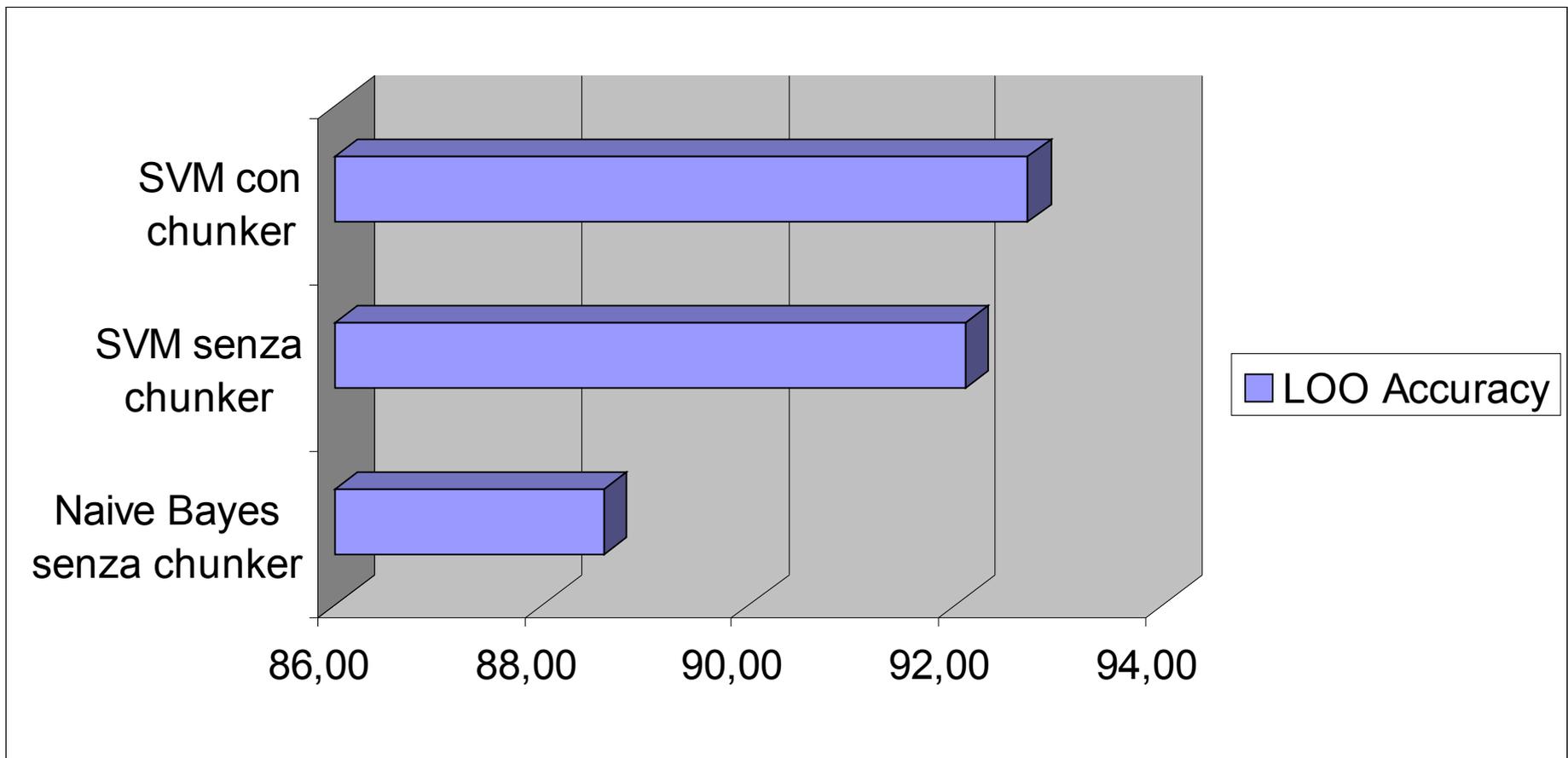
- Utilizzo di tutto il testo contenuto nelle disposizioni;

Train accuracy	Loo accuracy	Prime n parole con max infogain
99.7%	92.4%	tutte
99.7%	92.4%	100

- Eliminazione del testo contenuto nelle virgolette

Train accuracy	Loo accuracy	Prime n parole con max infogain
99.7%	92.7%	tutte
99.7%	92.7%	500

Confronto fra le tecniche di classificazione



Trattamento della semantica: Estrattore degli argomenti

1. *Il titolare che intenda procedere ad un trattamento di dati personali soggetto al campo di applicazione della presente legge è tenuto a darne notificazione al Garante.*

**Classificatore delle
Disposizioni**

Tipo di disposizione: "Obbligo"

...

Grammatica
dell'"Obbligo"

Grammatica della
"Sostituzione"

Grammatica della
"Definizione"

...

**Estrattore degli
Argomenti**

Argomenti:

Soggetto:

"Titolare"

Azione:

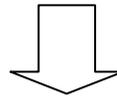
"Notificazione"

Controparte:

"Garante"

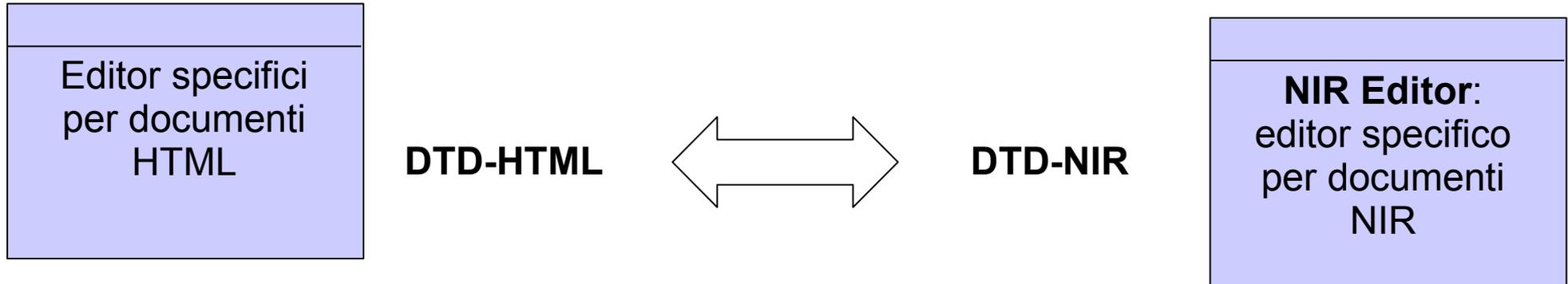
NREditor: **analisi dello scenario**

- Le DTD NIR individuano un sottinsieme ampio e complesso di documenti;
- La redazione di nuovi documenti o la conversione in formato XML-NIR di documenti esistenti, può essere un'attività complessa;
- Gli editor XML generalmente hanno molte limitazioni nelle funzioni di editing specifiche di una particolare classe di documenti:
 - Gestione delle partizioni e delle URN;
 - Implementazione di regole di tecnica legislativa;



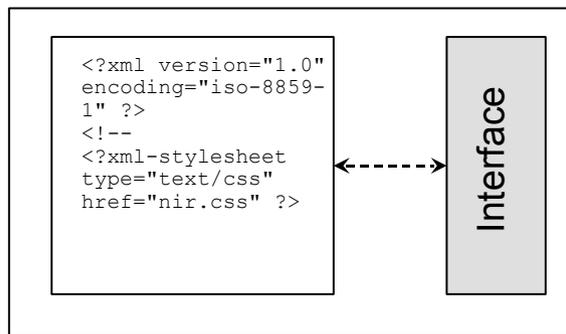
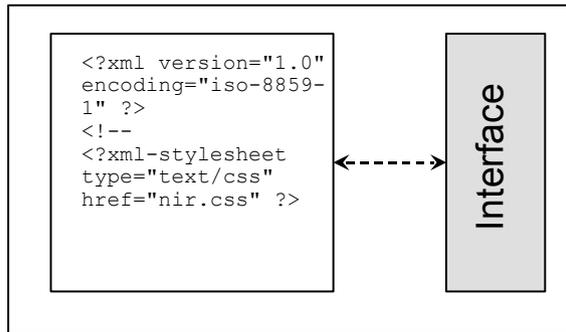
NREditor : ambiente specifico di redazione visuale per la gestione di documenti XML-NIR.

Criteri progettuali

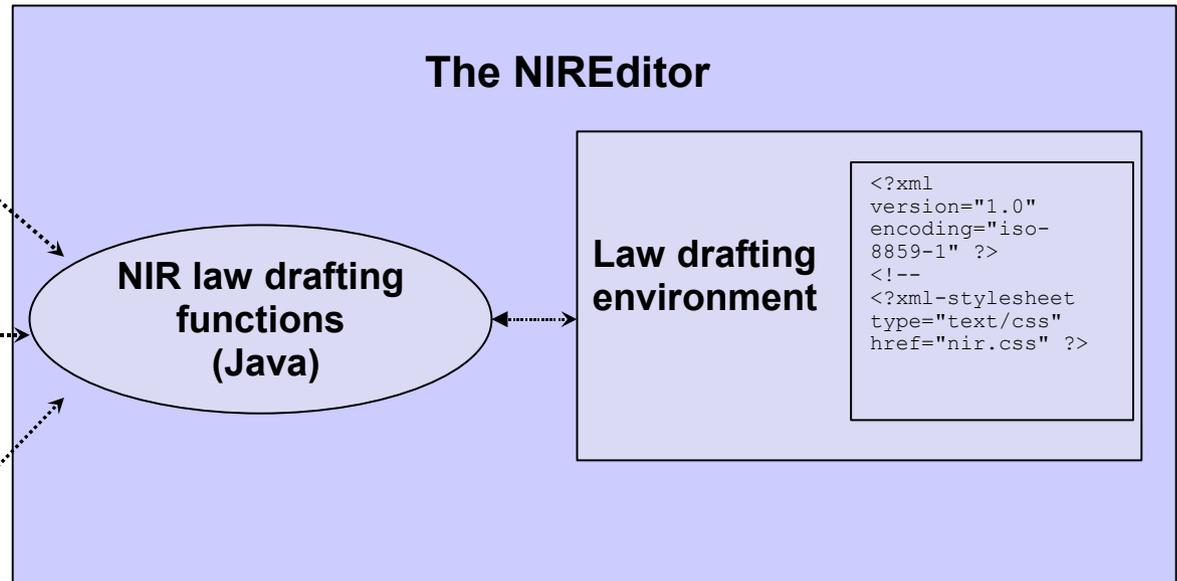


- NIR Editor:
 - Ambiente di redazione per il drafting legislativo;
 - Guida per l'utente alla corretta applicazione degli standard;
 - Ambiente di redazione specifico che consente la produzione di documenti XML-NIR validi.

General purpose XML editors



The NREditor architecture



- Nucleo di funzioni Java specifiche per NIR, integrate nell'ambiente per il drafting legislativo;
- Possono essere integrate in altri editor XML commerciali che supportano una Java API;

Caratteristiche di NREditor

- NREditor può gestire:
 - La struttura di un testo di legge (*profilo formale*);
 - La semantica di un testo di legge (*profilo funzionale*);
 - Documenti preesistenti da trasformare secondo gli standard XML-NIRusando funzionalità manuali e automatiche.
- Esegue solo operazioni valide secondo le DTD NIR (non è necessaria alcuna validazione);
- Le regole sono ottenute direttamente dalla DTD



Può essere facilmente adattato al altre DTD o a modifiche delle DTD NIR.

Ambiente di Redazione

- Finestra principale di drafting:
 - Rinumerazione automatica delle partizioni;
 - Costruzioni di riferimenti interni ed esterni;
 - Inserimento e cambiamento di posizione delle partizioni;
- Finestra secondaria:
 - Gestione del “plain text”;
 - Decomposizione e ricomposizione di sotto-strutture XML;
- Finestra per la gestione degli attributi degli elementi XML;
- Menù contestuale che suggerisce le azioni che possono essere eseguite in funzione della posizione del cursore.

Organizzazione di testi nuovi

- **Strategia formale**
 - Considera il testo dal punto di vista *formale*, ovvero composto da partizioni (collezioni di commi, articoli, ecc.);
 - Le partizioni da raggruppare e organizzare sono scelte direttamente dal redattore che indicherà quelle dello stesso rango da raggruppare insieme;
- **Strategia funzionale**
 - Considera il testo dal punto di vista *funzionale*: la componente elementare del testo è una disposizione (frammento regolativo);
 - Le disposizioni da raggruppare insieme sono scelte in base al loro contenuto, affinità, ecc.

Conclusioni e Sviluppi Futuri

- Panoramica sugli standard NIR:
 - URN
 - XML (Elementi formali e Metadati);
- *NIREditor*: un editor visuale capace di:
 - Convertire testi normativi preesistenti;
 - Produrre nuovi testi di normativi secondo gli standard NIR.
- *NIREditor* è in fase di test presso il Parlamento;