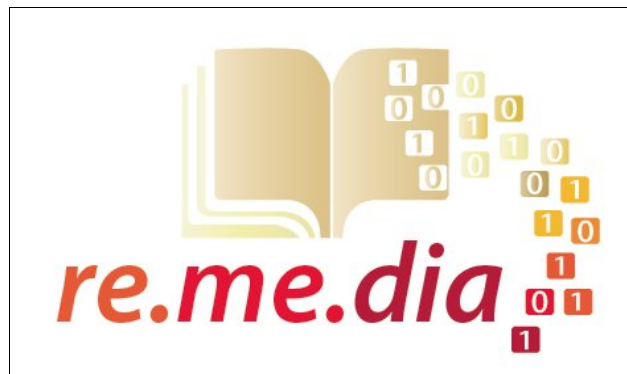




**Relazione sull'attività svolta**  
Periodo dic.2009-giu.2010

---

## 1 – Studio e realizzazione del logotipo “Re.Me.Dia”



Lo studio iniziale per la realizzazione del logotipo **Re.me.dia**. è partito dall’analisi delle finalità del progetto stesso: rendere fruibile, in formato cartaceo e digitale, la documentazione che va dal IX al XIII secolo conservata presso l’Archivio Capitolare di Arezzo concernente il “passato” di Arezzo e della Toscana sud orientale.

A questo punto si sono ricercati simboli ed elementi grafici che il più possibile potessero sintetizzare, in modo semplice, il maggior numero di significati inerenti gli obiettivi del progetto; infatti un logo è tanto più efficace quanto più riesce a trasmettere informazioni visive e concettuali di immediata comprensione:

“documenti, antico, restauro, trasformazione, digitalizzazione, fruibilità” queste le parole chiave ed i concetti che abbiamo mentalmente fissato e cercato di sintetizzare visivamente.

Da qui la stilizzazione dell'elemento grafico “libro”, che richiamasse l’idea delle fonti storiche e della cultura e i “bit” (unità elementare dell’informazione trattata da un elaboratore) per rappresentare la fase di digitalizzazione e della facile fruibilità.

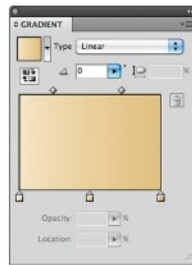
Il colore dell'elemento centrale è costituito da un gradiente di toni sul giallo che richiama l’aspetto dei vecchi documenti. Gli elementi sono correlati tra loro, cercando di esprimere il concetto di trasformazione o meglio di *digitalizzazione* e quindi la pagina destra del libro si divide in piccoli bit che assumono i valori “1” o “0” del codice binario.

Anche il gradiente dell’acronimo di “Restauro, Memoria, Digitalizzazione avanzata” vuole esprimere un concetto dinamico di trasformazione passando dall’arancio chiaro al rosso che nell'intensità ricorda l'inchiostro utilizzato per i Codici Miniati.



CMYK

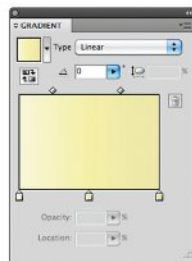
---



C4 M9 Y23 K0

C8 M16 Y42 K0

C12 M21 Y56 K0



C4 M9 Y23 K0

C8 M16 Y42 K0

C12 M21 Y56 K0



C4 M9 Y23 K0



C4 M9 Y23 K0

re.<sup>1</sup>

C4 M9 Y23 K0

me.<sup>0</sup>

C4 M9 Y23 K0

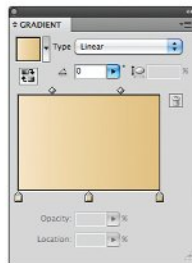
dia<sub>1</sub>

C4 M9 Y23 K0



RGB

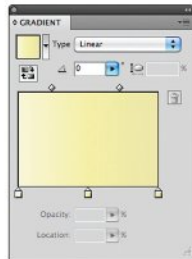
---



R242 G226 B197

R233 G207 B156

R225 G194 B130



R246 G241 B212

R246 G236 B165

R237 G228 B168



R253 G139 B31



R246 G241 B212

re.<sup>1</sup>

R241 G90 B34

me.<sup>0</sup>

R237 G28 B36

dia<sub>1</sub>

R190 G30 B45

## 2 - La fase di digitalizzazione

Il primo problema affrontato all'interno di Re.me.dia. è quello dell'applicazione di una tecnologia di riproduzione digitale (scansione) che consenta:

- fedeltà di riproduzione dei colori e dei materiali originali (pergamene in alcuni casi vecchie di mille anni);
- rispetto per il materiale e la forma originali. Le pergamene si presentano per lo più arrotolate e molte sono rimaste in quello stato per decenni;
- assenza di tecniche di scansione "invasive" che provochino alterazioni dello stato "naturale" della pergamena. Gli originali non possono essere scaldati, illuminati da fonti di luce diretta e compressi o piegati in forme che non siano quelle attuali;
- correzione degli errori e delle distorsioni dovute ad oggetti da scansionare non perfettamente in piano, con dislivelli anche di 3-4 cm.

Le tecniche tradizionali di fotografia digitale presentano numerose delle limitazioni ora elencate: per riprodurre antiche pergamene occorre illuminazione artificiale con fari di discreta potenza, che causano surriscaldamento e alterazione dell'ambiente originale di conservazione. La risoluzione, anche con macchine fotografiche digitali a 12 megapixel

non è adeguata, perché gli oggetti raggiungono anche i 90 cm. sul lato più lungo e la distanza tra l'obiettivo o l'oggetto deve essere aumentata per fornire una visione d'insieme. Viceversa, il ricorso ad obiettivi e focali di tipo "macro" consente di ottenere ottimi dettagli ma non una visione d'insieme della pergamena.

L'uso dello scanner piano (A4 o A3) è ugualmente sconsigliato poiché se da un lato risoluzione e dimensioni della scansione sono adeguate, dall'altro la pergamena è sottoposta a notevole

"schiacciamento" e pressione del coperchio dello scanner.

La soluzione ottimale è, in questi casi, l'utilizzo di scanner cosiddetto "planetario" (nella foto) che è dotato di particolari caratteristiche:

- La scansione avviene dall'alto, con un gruppo ottico che scorre lungo un binario mentre l'oggetto da scansionare rimane fermo sul piano di appoggio più in basso
- il piano è basculante e grazie alle due metà semovibili e inclinabili consente di regolare al meglio le distorsioni dovute alla pergamena non perfettamente in piano
- l'ottica prevede un meccanismo di correzione dell'immagine che elimina ombre, disallineamenti e curvature del foglio

La prima fase del progetto ha richiesto dunque la messa a punto di un procedimento di



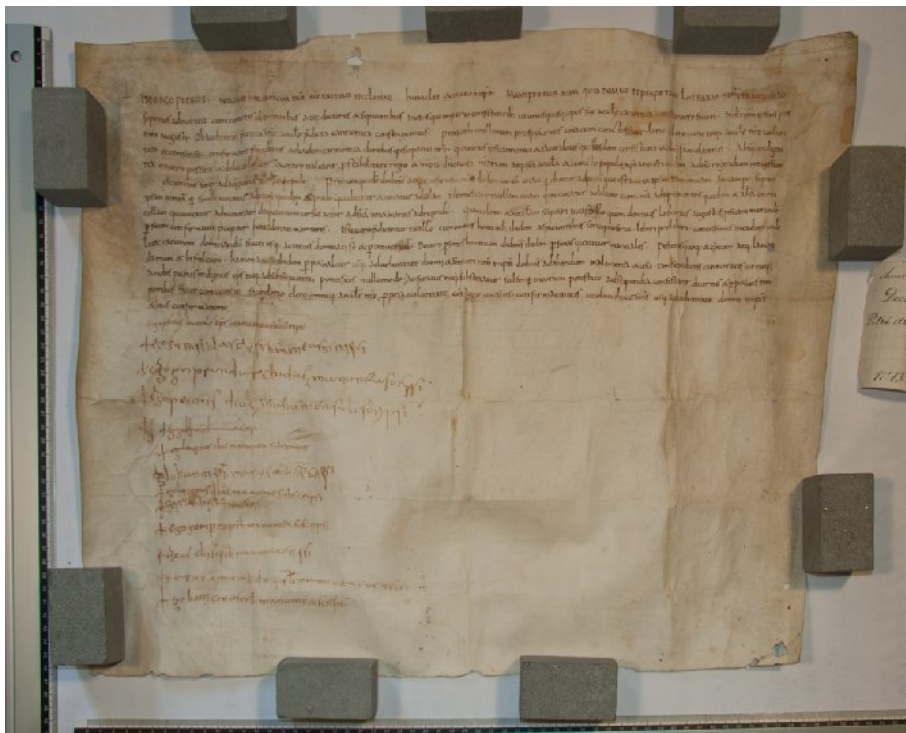
acquisizione digitale dell'immagine che garantisca una buona qualità media per esemplari di diverse dimensioni e in diverso stato di conservazione.

Come primo campione da studiare, i test sono stati effettuati su 22 pergamene, con la seguente segnatura nel catalogo dell'Archivio Diocesano di Arezzo:

- Canonica, numeri 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 20bis, 21, 23, 24, 25, 26, 27, 28 e 29
- Badia, numero 1

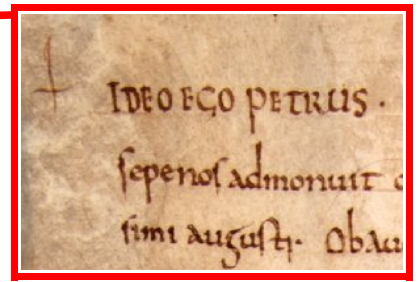
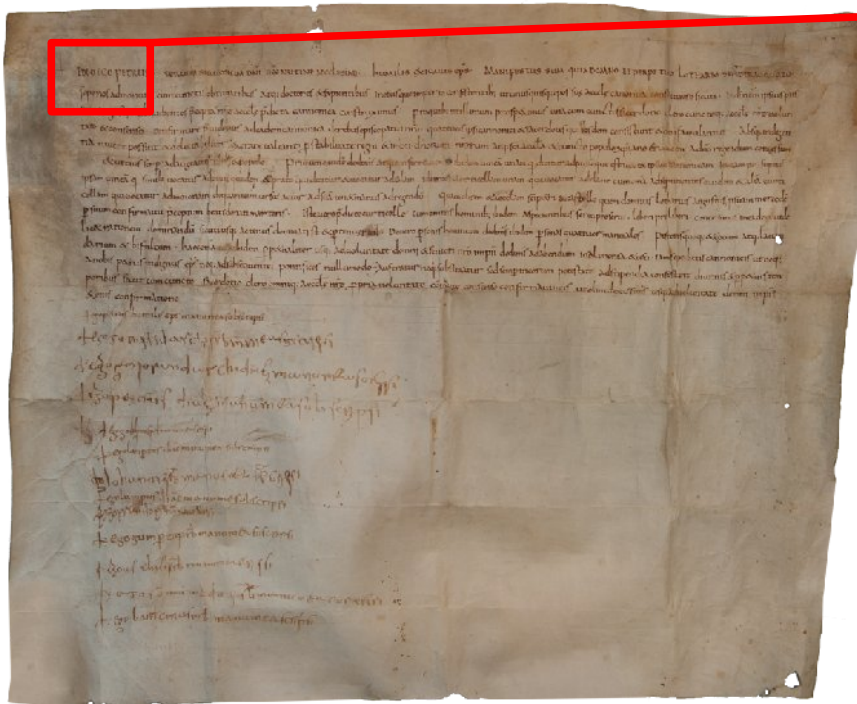
Per ogni immagine (*recto* e *verso*), oltre all'originale di scansione (formato TIFF a 300 dpi) sono stati prodotti *files* nei diversi formati jpeg dopo operazioni di fotoritocco con software Photoshop di Adobe. Il fotoritocco si è reso necessario per migliorare la fedeltà cromatica della riproduzione e l'eliminazione dei pesi in pietra utilizzati per tenere stesa la pergamena sul piano dello scanner. Le immagini successive mostrano un esempio tratto da *Canonica 13*, nella sua forma originale, in un particolare ingrandito e nella versione finale gestita dal software di marcatura tramite plug-in Zoomify (vedere caratteristiche del software di marcatura).

**Canonica 13, pergamena, originale, cm. 46X55 con righelli di misurazione e pesi in pietra**



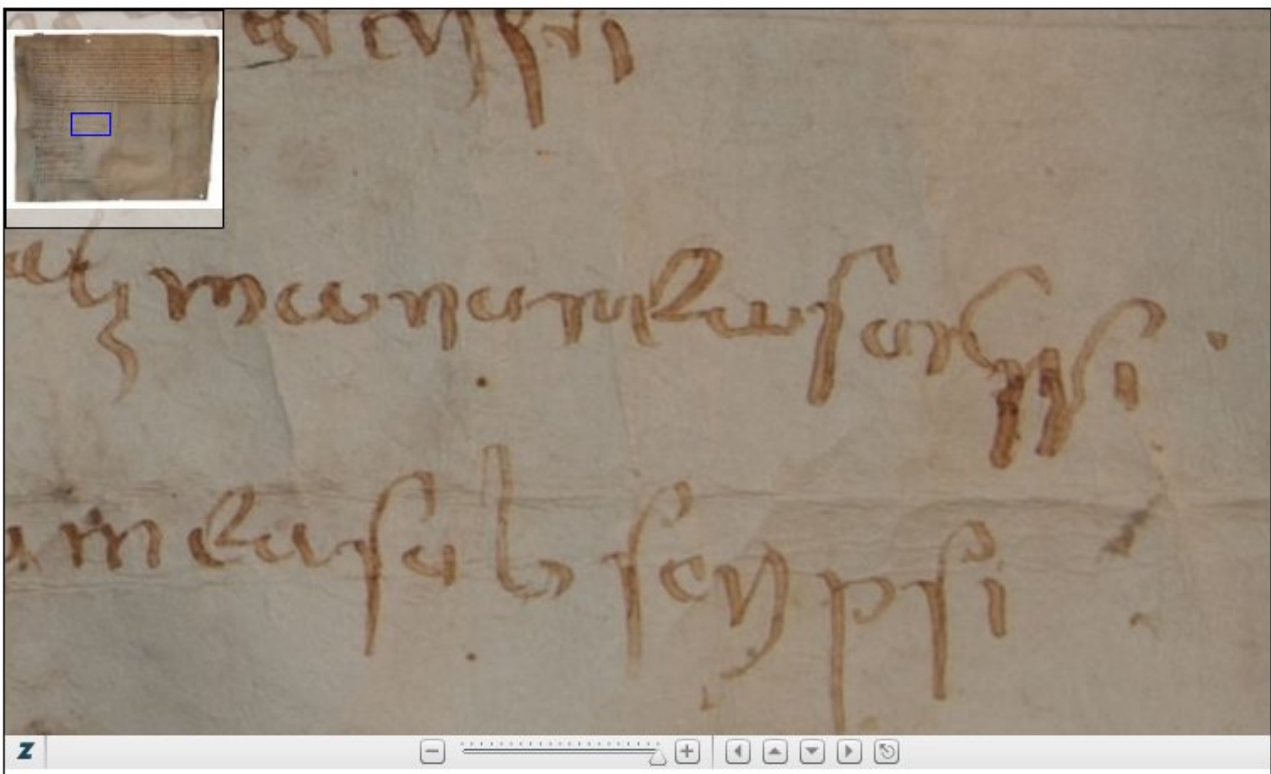
**Canonica 13, versione ritoccata/ripulita e ingrandita al 100%**





5,31 x 3,65 cm.

Canonica 13, particolare dell'immagine visibile, gestito dal software di catalogazione e marcatura



## 2 - Codifica dei testi secondo il modello TEI (Text Encoding Initiative)

La condivisione dei risultati di digitalizzazione e trascrizione dell'archivio pone successivamente il problema dell'adozione di uno standard che permetta una diffusione dei testi sul web in un formato leggibile da ogni sistema informatico esistente (in letteratura: *machine readable form*). E' questo un quesito chiave di quel ramo dell'informatica denominato "informatica umanistica" che propone l'applicazione delle tecnologie elettroniche alla gestione dei dati testuali e delle fonti, a oggi essenzialmente disponibili su supporto cartaceo.

L'iniziativa TEI (Text Encoding Initiative) è un progetto internazionale, attivo dal 1987 e promosso attualmente dalle tre maggiori associazioni mondiali di studiosi di scienze umane attraverso metodologie informatiche, la *Association for Computers and the Humanities* (ACH) la *Association for Computational Linguistics* (ACL) e la *Association for Literary and Linguistic Computing* (ALLC). Lo schema TEI è il risultato finale degli sforzi ventennali di definizione di uno standard internazionale che garantisce la validità scientifica della codifica, l'intercambiabilità e la portabilità dei testi in formato elettronico. TEI si occupa della creazione di un formato comune per la codifica di testi letterari in formato digitale attraverso l'utilizzo del linguaggio XML (eXtensible Markup Language), universalmente riconosciuto, grazie alla sua flessibilità ed estendibilità, come il miglior sistema per creare e descrivere basi di dati. L'obiettivo principale di TEI è la compatibilità dei testi codificati secondo questo standard e il progetto ha rilasciato negli anni differenti versioni, caratterizzate da modelli sempre più sofisticati e finalizzati alla trascrizione di *corpus* appartenenti a diverse discipline (teatro, prosa, poesia, dialoghi ecc.). La versione attualmente in uso, la TEI P5 prevede l'utilizzo di un nucleo comune a tutti i documenti TEI (TEI *core*) e moduli specifici per tipologie di testi, come il *TEI for manuscripts*, che prevede una serie di tag (marcatori) specifici per la descrizione del manoscritto antico.

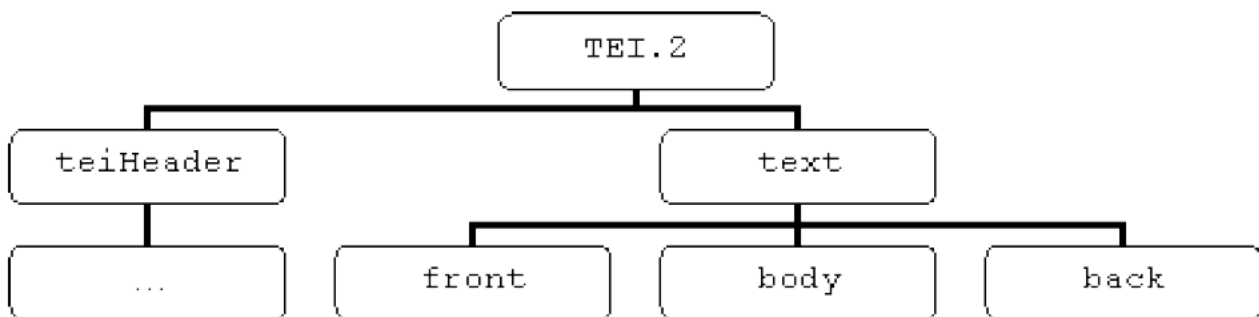
La scelta del modello TEI per la codifica elettronica dei testi permette di beneficiare dell'esperienza accumulata dall'azienda capofila, Tiphys S.r.l., nel campo dei linguaggi di programmazione per il web. Tiphys utilizza quotidianamente, per il proprio *core business*, i linguaggi XHTML e XML per la creazione di pagine web e l'interfacciamento con basi di dati su server remoti. Uno dei soci, Marco Marcellini, è inoltre docente di "linguaggi di marcatura dei testi" presso l'Università di Siena, Facoltà di Lettere e Filosofia di Arezzo. Il corso, attivato per l'a.a. 2009-2010, tratta l'uso del computer negli studi letterari e le edizioni digitali per le fonti documentarie.

Trascrivere un testo secondo il modello TEI significa andare oltre la sequenza di parole strettamente appartenenti al testo da trattare, sia esso la lista della spesa o una pagina della *Commedia* dantesca. Così come note a piè di pagina, titoli e sottotitoli costituiscono un



necessario complemento alla comprensione di un testo letterario, anche le informazioni contenute in un documento ogni testo codificato conformemente alle specifiche della TEI è costituito da due parti: – un TEI header (codificato con l'elemento <teiHeader>), contenente le informazioni editoriali concernenti il documento elettronico, – e un TEI text (codificato con l'elemento <text>) contenente la trascrizione codificata del testo in versione integrale. – Ogni TEI header consta di quattro parti, una delle quali deve necessariamente ricorrere in ogni testo codificato secondo le raccomandazioni della TEI, ed è l'elemento <fileDesc> contenente i METADATI, mentre le altre tre, <encodingDesc>, <profileDesc>, <revisionDesc>, sono opzionali – L'elemento text si divide, a sua volta, in quattro elementi: – <front> (opzionale), <body> (obbligatorio), <group> (obbligatorio) e <back> (opzionale).

Lo schema sottostante sintetizza la composizione di un testo codificato secondo TEI:



Basando il modello di codifica su TEI P5 for manuscript<sup>1</sup>, set di marcatori TEI comprensivo del TEI CORE (nucleo comune a tutti i documenti TEI) e di specifici tags per la descrizione del manoscritto e delle fonti documentarie, così potrebbe apparire l'inizio del file XML corrispondente ad un primo esperimento di codifica.

```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader type="ISBD-ER">
    <fileDesc>
      <titleStmt>
        <title type="main">Codice Diplomatico Aretino</title>
        <author>Marta Calleri</author>
        <funder>Universita di Siena</funder>
      </titleStmt>
      <editionStmt>
        <edition>Versione 0.9b</edition>
      </editionStmt>
    </fileDesc>
  </teiHeader>
  <text>
    <front>
      <p>...</p>
    </front>
    <body>
      <p>...</p>
    </body>
    <back>
      <p>...</p>
    </back>
  </text>
</TEI>
  
```

<sup>1</sup> - La descrizione di questo specifico set di tag è online all'uri <http://www.tei-c.org/Guidelines/Customization/>

```

<extent>
  <seg type="designation">Codice Diplomatico Aretino</seg>
    <seg type="size">21 files : 3 megabytes dimensione
provvisoria</seg>
</extent>
<publicationStmt>
  <authority>Progetto Re.Me.Dia.<date>2010-06-08</date></authority>
  <distributor>
    <orgName type="UNISI">Universita di Siena</orgName>
    <placeName>Sede di Arezzo</placeName>
    <address>
      <orgName type="CISLAB">C.I.S.L.A.B. - Centro
Interdipartimentale di Studi sui Beni Librari e Archivistici</orgName>
      <orgName type="LETTEREAREZZO">Facolta di lettere e Filosofia
di Arezzo</orgName>
      <addrLine>Viale Cittadini 33</addrLine>
      <addrLine>Arezzo</addrLine>
      <addrLine>52100</addrLine>
    </address>
    <email>info@remedialab.it</email>
  </distributor>
  <idno type="UNISI">0011</idno>
  <availability n="0001s" status="free" rend="visible">
    <p>Freely available, is a sample</p>
  </availability>
  <date>2010-06-08</date>
</publicationStmt>
<notesStmt>
  <note>Documento di prova, test n.01</note>
  <note>Canonica 1</note>
  <note type="summary">Sigizo del fu Petrone vende al presbitero
Leone del fu Domenico la metà di un appezzamento di terra con casa e vigna
situato nella pieve di San Pietro in Agello, in casale Paterno, in
località Givestelle, al prezzo di 20 soldi.</note>
  ... ..
</TEI>

```

Il documento complessivo, corrispondente alla trascrizione della pergamena catalogata come "Canonica 1", sarà un file .XML di circa un migliaio di righe in linea con le direttive TEI P5.

### 3 - Caratteristiche del software

Il passaggio dalla semplice trascrizione del documento, realizzata con programmi di videoscrittura come Microsoft Word, all'edizione digitale ovvero al file XML costruito rispettando le regole TEI è processo difficile e suscettibile di numerosi errori.

Esistono in commercio software come Oxygen ([www.oxygenxml.com](http://www.oxygenxml.com)), che nascono come editor xml generici e quindi pensati per programmatori o sistemisti piuttosto che per studiosi di documenti antichi. Il loro utilizzo richiede conoscenze informatiche avanzate e manca del tutto il collegamento visivo con la fotografia digitale del documento da marcare.

L'idea di base è quindi quella di realizzare un software di ausilio alla trascrizione e marcatura delle pergamene, dotato delle seguenti caratteristiche:

- aumento della facilità d'uso
- collegamento con le immagini delle pergamene, con funzionalità di zoom
- interfaccia con meno funzioni ma adatta ad utenti estranei alla struttura tipica di un software usato da programmatori e tecnici
- automazione di quante più funzioni possibile. Marcando documenti dello stesso tipo (medesima epoca, luogo, fondo, lingua) deve essere possibile impostare una sola volta le informazioni di base sulle caratteristiche. In pratica, il blocco TEI HEADER deve essere gestito in automatico dal software
- anche se deve essere garantita l'esportazione in formato XML TEI, i dati del singolo documento vengono memorizzati PRIMA in un database MySQL, per poter disporre con maggior facilità di indici e ricerche. Alternativamente, i soli file xml costringerebbero ad implementare un software aggiuntivo per la ricerca e l'indicizzazione dei medesimi
- utilizzo in rete, con una soluzione web-based
- rilascio sotto licenza open-source

Lo schema sottostante riassume l'architettura pensata per il software, che abbiamo scelto di chiamare CDAeDit (editor xml del codice diplomatico aretino digitale). Seguono alcune schermate esemplificative della prima versione del software.

Figura 1 – Architettura del software CDAeDit

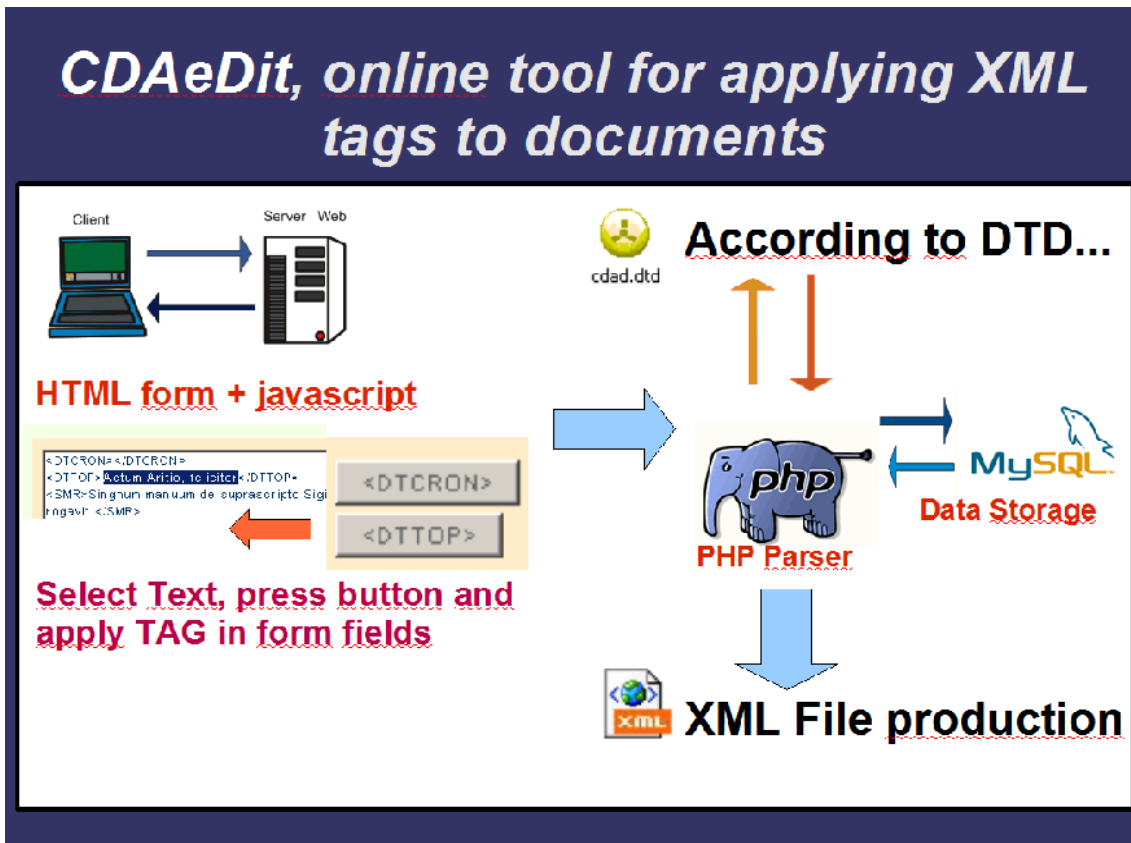


Figura 2 – Accesso all'elenco dei documenti codificati, con indicazione dell'autore e dello stato



Figura 3 – La descrizione del manoscritto e i campi TEI HEADER sono affidati a campi fissi

The screenshot shows the 'DOCUMENT EDITOR' interface. It contains two main sections:

- General information on file, number, collection, dates:** A grid of input fields for:
  - NAME: Sigizo figlio del fu Patrone...
  - TIT. DOC.: Documento di prova
  - USER: marco
  - AREA: Codice Diplomatico Aretino digitale
  - FONDO: Archivio Santa Maria in Gradi di Arezzo
  - NUMERO: 1
  - FILE XML: prova1.xml
  - DATA: 1029, giugno, Arezzo
- REGESTO, APPARATO and other info on edition, aspect and status:**
  - REGESTO:** A text area containing a summary of the document's content.
  - TRADITIO:** Originali [A]
  - NOTITIA:** (Empty field)
  - AUTII:** Marta Calcei
  - EDIZIONE:** Edizione di prova
  - BIBLIO:** (Empty field)
  - OSSERVAZIONI:** Pergamena di formato regolare. Ca. cm. 20 x 44,5. La pergamena presenta macchie di umidità lungo il margine destro, che non pregiudicano la lettura del testo. Scrittura disposta secondo il lato corto.

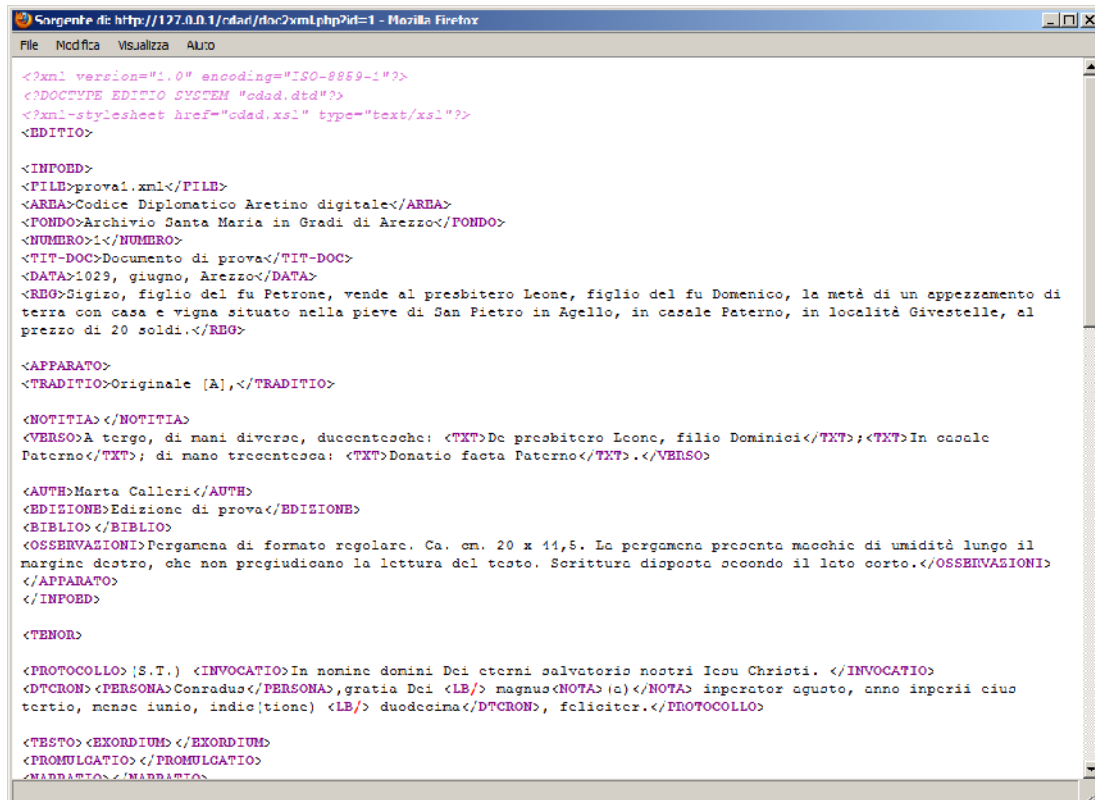
Figura 4 – La trascrizione del testo si avvale di pulsanti che contengono i marcatori fondamentali per ogni parte del documento. Questo evita errori di annidamento e tag non consentiti.

The screenshot shows the 'PROTOCOLLO' and 'TESTO' sections of the editor. The 'TESTO' section contains a large text area with Latin text and TEI tags. To the right of the text area is a vertical toolbar with buttons for various TEI tags:

- <INVOCATIO>
- <DTCRON>
- <PERSONA>
- <NOTA>
- <LB/>
- <EXORDIUM>
- <PRIMUM>
- <NARRATIO>
- <DISPOSITIO>
- <AUCT>
- <RECIP>
- <RES>
- <FORMIUM>
- <CLAUSULA>
- <PERSONA>
- <TOP>
- <MCRTOP>
- <NOTA>
- <LD>

The text in the 'TESTO' section includes tags like <EXORDIUM>, <PRIMUM>, <NARRATIO>, <DISPOSITIO>, <AUCT>, <RECIP>, <RES>, <FORMIUM>, <CLAUSULA>, <PERSONA>, <TOP>, <MCRTOP>, <NOTA>, and <LD>.

Figura 5 – Il risultato finale è un file XML che negli obiettivi del progetto deve essere validato DTD TEI



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<DOCTYPE EDITIO SYSTEM "edad.dtd"?>
<xml-stylesheet href="cdad.xsl" type="text/xsl"?>
<EDITIO>

<INFOED>
<FILE>prova1.xml</FILE>
<AREA>Codice Diplomatico Aretino digitale</AREA>
<PONDO>Archivio Santa Maria in Gradi di Arezzo</PONDO>
<NUMERO>1</NUMERO>
<FIT-DOC>Documento di prova</FIT-DOC>
<DATA>1029, giugno, Arezzo</DATA>
<RBG>Sigizo, figlio del fu Petrone, vende al presbitero Leone, figlio del fu Domenico, la metà di un appezzamento di
terra con casa e vigna situato nella pieve di San Pietro in Agello, in casale Paterno, in località Givestelle, al
prezzo di 20 soldi.</RBG>

<APPARATO>
<TRADITIO>Originale [A],</TRADITIO>

<NOTITIA></NOTITIA>
<VERSO>A tergo, di mani diverse, duecentesche: <TXT>De presbitero Leone, filio Dominici</TXT>; <TXT>In casale
Paterno</TXT>; di mano trecentesca: <TXT>Donatio facta Paterno</TXT>.</VERSO>

<AUTH>Marta Calleri</AUTH>
<EDIZIONE>Edizione di prova</EDIZIONE>
<BIBLIO></BIBLIO>
<OSSERVAZIONI>Pergamena di formato regolare. Ca. cm. 20 x 14,5. La pergamena presenta macchie di umidità lungo il
margine destro, che non pregiudicano la lettura del testo. Scrittura disposta secondo il lato corto.</OSSERVAZIONI>
</APPARATO>
</INFOED>

<TENOR>

<PROTOCOLLO>[S.T.] <INVOCATIO>In nomine domini Dei eterni salvatoris nostri Iesu Christi. </INVOCATIO>
<DTCRON><PERSONA>Conradus</PERSONA>, gratia Dei <LB/> magnus<NOTA> (S)</NOTA> imperator augustus, anno imperii eius
tertio, mense iunio, indictione <LB/> duodecima</DTCRON>, feliciter.</PROTOCOLLO>

<TESTO><EXORDIUM></EXORDIUM>
<PROMULGATIO></PROMULGATIO>
<NARRATIO></NARRATIO>
```

La sperimentazione sul software è ovviamente ancora in corso. In questa prima fase del progetto vi sono ancora obiettivi da raggiungere:

- maggior facilità d'uso
- stabilità e miglioramento dei margini di tolleranza degli errori commessi dall'utente
- miglior collegamento con la TEI P5, per ora limitata ad una sola specifica *Document Type Definition (TEI for manuscripts)*
- miglioramento dell'interfaccia grafica, anche grazie al ricorso di AJAX come evoluzione alle funzioni dei pulsanti ora gestite con il semplice JAVASCRIPT

Nei prossimi mesi il software verrà sviluppato e reso produttivo, passando dall'attuale release *alpha*, dove i meccanismi di base sono appena abbozzati, ad una versione stabile di produzione, rilasciata sotto licenza GNU-GPL (open source).





## Progetto Re.Me.Dia.

### Rendicontazione seconda fase

luglio 2010-marzo 2011

### ***Introduzione e grado di raggiungimento degli obiettivi***

Terminata la progettazione del database e dell'infrastruttura generale del software per l'ausilio alla trascrizione delle pergamene del Codice Diplomatico Aretino, la seconda fase del progetto ha visto impegnata la nostra azienda su un duplice fronte: da un lato è stato necessario concertare, con il comitato scientifico, un modello di documento scritto in linguaggio XML che fosse contemporaneamente corretto da un punto di vista sintattico (well-formed, secondo la sintassi imposta dal manuale TEI-P5<sup>1</sup>) e che contenesse tutta una serie di TAG atti alla descrizione completa dei manoscritti/pergamene del periodo storico considerato. In pratica si è trattato di elaborare un modello concettuale che fosse validato dalle seguenti fasi operative:

- studio della descrizione del manoscritto;
- stesura di un elenco condiviso di elementi, informazioni e caratteristiche tratte dall'edizione cartacea del Codice Diplomatico Aretino;

---

1 TEI o Text Encoding Initiative è un progetto ventennale, riconosciuto a livello internazionale, che propone una serie di regole scritte in linguaggio XML per la trascrizione di qualsiasi tipo di documento, dalla lista della spesa ai rotoli del Mar Morto. Il progetto, ben documentato al sito [www.tei-c.org](http://www.tei-c.org) è stato già citato più volte nella prima parte di Re.Me.Dia. e ci pare superfluo descriverlo ulteriormente. Ci pare sufficiente qui ricordare che le regole sintattiche proposte dalla versione 5 del manuale TEI sono quelle adottate come standard per l'edizione digitale del nostro progetto e che in **appendice 3** è possibile trovare materiale didattico, utilizzato per la formazione interna dei partecipanti al progetto, che illustra in maniera sintetica le caratteristiche di un documento XML scritto secondo lo standard TEI.

- traduzione di tali elementi in TAG xml accettati e sintatticamente corretti dal modello TEI;
- elaborazione di un modello XML-TEI sintatticamente corretto che dovrà risultare come prodotto finale del software di marcatura
- creazione dell'interfaccia di base del software di ausilio alla marcatura, con definizione dei meccanismi di lettura/scrittura nel database e della gestione delle immagini delle pergamene
- sperimentazione delle prime criticità del software

Volendo tradurre il flusso di lavoro in termini di obiettivi operativi, possiamo affermare che questa seconda fase del progetto giunge fino a parziale raggiungimento dell'obiettivo 3, che prevedeva la pubblicazione dell'edizione cartacea delle trascrizioni delle pergamene appartenenti al Codice Diplomatico Aretino e in contemporanea la loro versione “elettronica”, pubblicazione gestita tramite il software di marcatura.

Le pagine successive, comprensive di n.3 appendici contenenti materiale di *reference* e didattico, descrivono le singole fasi operative e i risultati ottenuti.

## **Confronto con altri modelli di marcatura**

Nell'elaborare il modello di marcatura dei documenti specifico per il progetto Re.Me.Dia. abbiamo ritenuto utile confrontare le nostre esperienze in tema di linguaggio XML e applicazione di TEI ai documenti antichi, con quelle di progetti già avviati in Italia e all'estero da parte di importanti istituzioni culturali internazionali.

L'occasione si è presentata nel mese di dicembre 2010 in occasione del seminario tenuto a Siena (collegio Santa Chiara), per la scuola di dottorato in “Scienze del Testo”, dal prof. Michele Ansani dell'Università di Pavia. Il docente, principale autore del “Codice Diplomatico Lombardia Medievale” (<http://cdlm.unipv.it/>) ha tenuto un ciclo di lezioni con il seguente calendario:

- 1 dicembre 2010 - *L'edizione digitale di fonti documentarie: i rapporti tra applicazioni informatiche e discipline medievistiche.*
- 2 dicembre - *L'edizione digitale di fonti documentarie: fonti storiche e informatica. I*

*modelli di edizione documentaria e la tradizione*

- 3 dicembre - Il ‘nodo’ della codifica testuale: il dibattito internazionale; illustrazione delle maggiori esperienze di edizione digitale e discussione dei modelli
- 4 dicembre - L’edizione digitale di fonti documentarie: la costruzione di uno schema di codifica per le fonti documentarie basato su XML
- 10 dicembre - L’edizione digitale di fonti documentarie: Laboratorio di codifica testuale applicata a fonti documentarie italiane dei secoli X-XII

L'altro progetto con cui ci siamo confrontati prima dell'elaborazione di un nostro modello XML per la marcatura del codice diplomatico è quello redatto dall'*École nationale des chartes* con sede a Parigi, ente universitario che è parte, assieme al King's College di Londra e al nostro Centro Interdipartimentale di Studi sui Beni Librari e Archivistici dell'Università di Siena, sede di Arezzo, di un gruppo di lavoro internazionale per l'edizione digitale dei documenti.

Ci preme sottolineare che entrambe le istituzioni ora citate hanno “pubblicato” sul web un loro codice diplomatico digitale. In nessun caso però, i documenti digitali xml sono stati prodotti mediante uno specifico software (si è utilizzato un editor standard), né vi è collegamento o inclusione nel database della foto-riproduzione della pergamena o del manoscritto originale; inoltre il codice diplomatico lombardo non si avvale dello schema TEI, di fatto rendendo i documenti prodotti non confrontabili con altri in quanto non aderenti ad uno standard riconosciuto.

L'occasione di confronto con il prof. Ansani e con quello che definirei “modello francese” sviluppato secondo TEI-P5 ci ha consentito però di “mettere sul tavolo” i dubbi relativi all'adozione di questo o quel *tag* (es. adozione di <DIV> piuttosto che <P<sup>2</sup>>) e di verificare sul campo la bontà dell'infrastruttura software da noi progettata, specie per quanto riguarda le funzionalità di affiancamento dell'immagine digitale al testo.

---

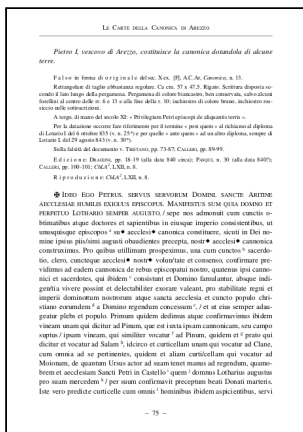
2 DIV e P sono utilizzati dal modello TEI-P5 come marcatori per la suddivisione del testo.

## Definizione dei campi necessari alla marcatura

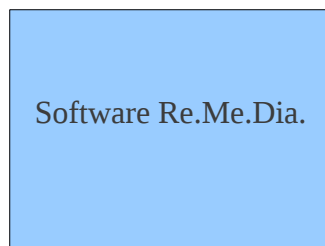
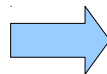
Dopo lo studio delle regole sintattiche proposte dallo standard TEI e il confronto con altre esperienze di settore è stato necessario elaborare, in stretta connessione con il comitato scientifico, un elenco di marcatori definibili come “fondamentali” per la descrizione e trascrizione dei manoscritti facenti parte il Codice Diplomatico Aretino.

L'appendice 1 riporta l'elenco ora descritto; si noterà che i marcatori sono suddivisi a gruppi, alcuni dei quali atti a descrivere le caratteristiche fisiche e generali del manoscritto, altri a rappresentare la suddivisione del documento in parti (formule di invocazione, saluto, formule legali ecc.) e altri ancora a gestire a livello di database elementi da indicizzare quali nomi, luoghi, cariche, date e personaggi storici. Tutte le voci citate derivano da una visione “scientifico-letterale” e non “informatica” dello schema in cui può essere ricondotto un antico manoscritto. Lo schema logico di riferimento è sempre il medesimo: dall'edizione cartacea all'edizione digitale del medesimo documento il passo deve essere breve, come raffigurato nello schema sottostante. Vi è uno strato intermedio di elaborazione, rappresentato dal software Re.Me.Dia. che ha appunto l'obiettivo di consentire una facile trasformazione, dall'edizione di un documento progettata per la stampa a quella digitale progettata per il web e conforme allo standard TEI.

### La trascrizione di un testo: dall'edizione cartacea a quella digitale attraverso il software Re.Me.Dia.



Edizione cartacea



**Punti di forza:**  
Non richiede conoscenza di linguaggio XML, né TEI

Affianca al testo l'immagine, così da consentire una più facile trascrizione

Gestisce in un database nomi, luoghi e date

Consente ricerche e indici



Edizione digitale = file XML

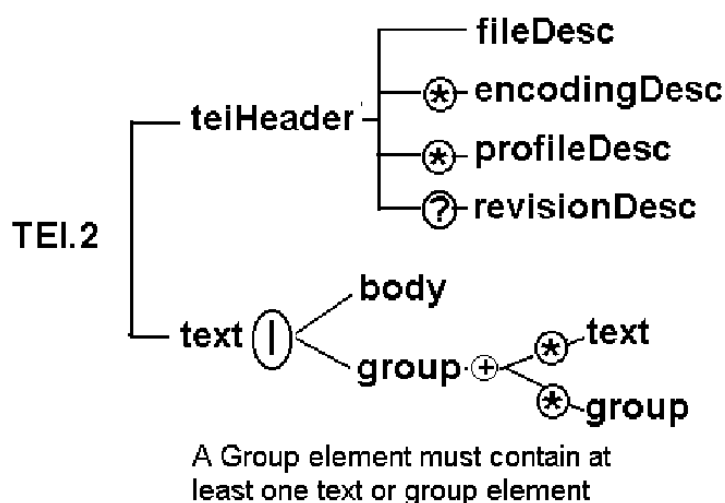
redatto secondo lo standard TEI

## Il modello XML-TEI specifico di Re.Me.Dia.

Definite le caratteristiche e i contenuti minimi da includere sotto forma di marcatori, il passo successivo è la definizione di un modello XML che da un lato contenga tutti gli elementi richiesti e che dall'altro, condizione necessaria, rispetti la grammatica e l'insieme di regole definite da TEI.

La figura che segue illustra lo schema di un documento TEI di base e la suddivisione del file in due parti fondamentali:

- il *teiHeader* (o testata del documento) che andrà a contenere la descrizione del manoscritto e tutte le meta-informazioni sulla pergamena originale
- il *text*, ovvero il corpo del documento che conterrà la trascrizione vera e propria del testo, arricchita di tutti quei meta-dati idonei alla descrizione di note, caratteri non comprensibili o abbreviazioni, nomi di persone, luoghi ecc.



Per la prima parte del documento (*teiHeader*) è stato tratto ampio spunto dal gruppo di studio all'interno del progetto TEI definito “TEI for manuscripts<sup>3</sup>” che fornisce un set di marcatori specifici per appunto per la descrizione delle fonti. In mancanza di letteratura specifica invece, per *text*, *body* e *group* si è invece elaborato un modello che fa largo uso del tag <DIV TYPE=... > per suddividere il manoscritto in blocchi corrispondenti alle varie formule di invocazione, dichiarazione, saluto e così via (*Invocatio*, *Apprecatio*, *Intitulatio*, *Inscriptio*, *Salutatio*,

<sup>3</sup> [www.tei-c.org/Activities/SIG/Manuscript/index.xml](http://www.tei-c.org/Activities/SIG/Manuscript/index.xml)

*Promulgatio, Narratio, Dispositio, Formule Clausule).*

Il risultato, applicato ai documenti Canonica 13 e Canonica 6 è riportato in **appendice 2**. Sarà ormai chiaro che il file XML che rappresenta l'edizione digitale del testo antico dovrà essere il prodotto finale del software di marcatura, oggetto principale del progetto.

## ***La prima versione (0.9a) del software di marcatura***

Conclusa la fase di definizione del modello, le attività si sono concentrate sulla realizzazione del software di marcatura. Il software è attualmente in fase di alfa-test ma le schermate che seguono ne mostrano già i meccanismi fondamentali:

- Architettura *web-based* che gira su sistema operativo Linux e che si avvale di *web-server* Apache e database MySQL.
- Dal lato utente, è sufficiente un normale *browser* web per l'accesso, la lettura e la modifica dei documenti.
- Il linguaggio di programmazione prescelto è un mix di PHP (per l'accesso al database) e Javascript per le funzioni interattive (pulsanti, zoom delle immagini). Le pagine sono costruite in HTML e CSS.
- L'accesso ad ogni documento è gestito tramite *tabs* che consentono di spostarsi rapidamente tra immagine/zoom, descrizione del documento e sua trascrizione.
- La trascrizione e marcatura avvengono tramite pulsanti che, analogamente ai consueti meccanismi usati in un word-processor per applicare la formattazione del testo, consentono di selezionare una o più parole e applicare con un solo clic il tag appropriato.



Fig. 1 – elenco dei file inseriti nel database, con indicazione dello stato di avanzamento della trascrizione

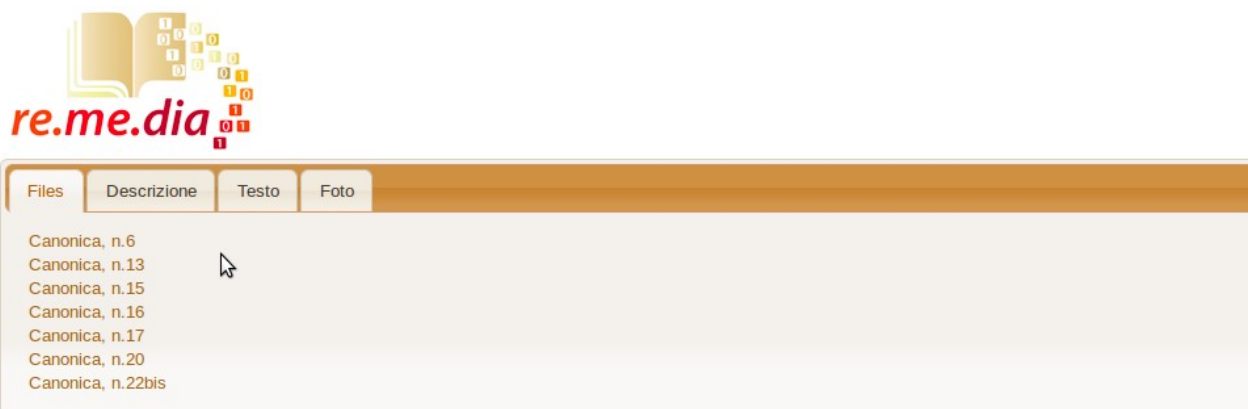
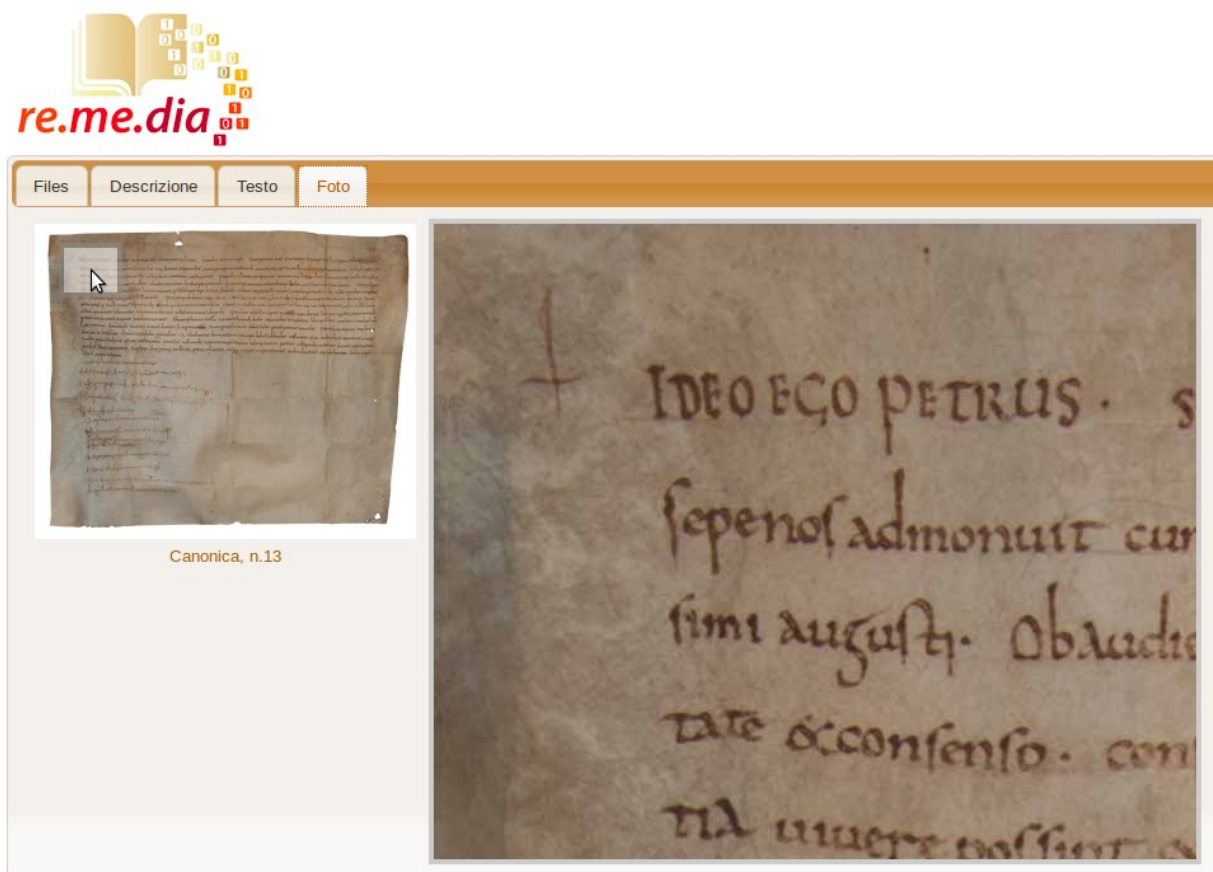


Fig. 2 – la riproduzione della pergamena consente lo zoom dinamico, gestito spostando il mouse sopra l'immagine



**Fig. 3 – lo zoom consente di spostarsi nelle varie parti del documento, evidenziando parti mancanti, parti abrase o lettere cancellate, ausilio fondamentale per la trascrizione. In alternativa si dovrebbe disporre del manoscritto originale o ricorrere a foto-riproduzioni su carta**



**Fig. 4 – La parte descrittiva del manoscritto è gestita separatamente dalla finestra di trascrizione**



**Fig. 5 – Nella finestra dedicata alla trascrizione del testo, i diversi marcatori possono essere applicati tramite selezione della porzione di testo e clic sul pulsante relativo**



Il software è ancora in fase di costruzione (sono stati testati solo i meccanismi di base) e il test sul campo avverrà in settembre con l'utilizzo sulle trascrizioni del primo volume del Codice Diplomatico Aretino. Il completamento del software, l'eliminazione dei bugs e la diffusione dei risultati sul sito [www.remedialab.it](http://www.remedialab.it) saranno oggetto della terza e ultima fase, da concludere entro dicembre 2011.

Il responsabile scientifico per Tiphys S.r.l.

Dott. Marco Marcellini



## ***APPENDICI***

## 1 - Definizione dei campi per la marcatura

Versione 1.0 approvata Comitato Scientifico novembre 2010

### Contenuti minimi da includere nella parte descrittiva del documento:

Numero documento

Data

Regesto (breve sintesi del contenuto)

Titolo che definisce la natura del documento (Diploma, Privilegium, Littera Gratiiosa, littera executoria, charta, breve, placitum, instrumentum, scritture originali, laudes ...) - **in corso di approvazione**

### Altre informazioni da includere nella descrizione del manoscritto (sezione TEI-HEADER del documento XML)

Note introduttive

Traditio

Archivio

Fondo;

Collocazione archivistica;

Descrizione fisica

Note dorsali (Notae, Notitiae, Rogationes, Dicta ...) / Processi di autentica .

Annotazioni tergalì

Osservazioni (spiegazioni datazione, rinvii ad altri docc. ... altro)

Bibliografia: Edizioni, Regesti, pubblicazioni che ricordano il documento (**in corso di approvazione**)

---

## SUDDIVISIONE DEL MANOSCRITTO IN PARTI ( **elemento <DIV>** )

### **Protocollo:**

#### **se documento privato:**

Invocatio

Data cronica / Data topica

Apprecatio

#### **se documenti pubblici:**

Invocatio

Intitulatio (Formulae humilitatis / pietatis?)

Inscriptio

Salutatio / Formula di perpetuità

### **Testo:**

Arenga

Promulgatio

Narratio

Dispositio

Formule

Clausule

Rinunce

Additum (ad es. le pattuizioni aggiuntive)

Res (per individuare elenchi e descrizione di beni oggetto di scambio)

Decretum per i Privilegi papali

Sanctio / Penale

Corroboratio / Rogatio



### **Escatocollo:**

Subscriptio regis / papae

Recognitio

Sottoscrizioni / Signa manuum di

Attori

Consensienti

Estimatori

Fideiussori

Testimoni

Giudici

### **Completio**

Data cronica / Data topica

## **NOMI DI PERSONA, LUOGHI, ISTITUZIONI ECCLESIASTICHE**

### **Codifica dei nomi di persona:**

Nome

Nome normalizzato

Identificazione

Professione di legge

Attributi di parentela: figlio, padre, madre, sorella, fratello, marito, moglie ecc.

Qualifiche: scrittore, redattore del documento, rogatario, autenticatore / scrittore, autenticatore / sottoscrittore, autore, destinatari, avvocati, arbitri, difensori, estimatori, giudici, fideiussori, testimoni, consenzienti, rinuncianti... ecc.

### **Codifica dei nomi di luogo:**

Nome

Nome normalizzato

Identificazione del toponimo

Microtoponimo

Nome normalizzato del microtoponimo

Identificazione del microtoponimo

### **Codifica delle istituzioni ecclesiastiche.**

Identificazione

Toponimo

### **Scioglimento delle abbreviazioni:**

Marcatura delle parentesi tonde limitate solo allo scioglimento dei casi dubbi

Marcatura delle parentesi quadre: restituzione di ciò che non è più leggibile

Marcatura del segno di fine riga

Marcatura dei Monogrammi, Rota, Bene Valetè ....

Marcatura delle litterae elongate

Marcatura degli eventuali spazi bianchi

Marcatura delle parentesi uncinata per eventuali omissioni di lettere, sillabe, parole ...

### **Apparato critico:**

Note

Note storiche

## **2 – Esempi di trascrizione XML-TEI di documenti del codice diplomatico aretino**

CANONICA 6 e CANONICA 13

Nota: i modelli XML-TEI proposti sono ancora in fase di validazione, ma contengono tutti gli elementi richiesti dal comitato scientifico, relativamente alla suddivisione del manoscritto, sua descrizione, datazione e definizione degli indici di nomi, luoghi, cariche ecclesiastiche.

```

<?xml version="1.0" encoding="UTF-8"?>
<?oxygen RNGSchema="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_ms.rng" type="xml"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">

  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title></title>
        <author>Carolus Magnus</author>
        <respStmt>
          <resp>Marta Calleri</resp>
          <name>Marta Calleri</name>
        </respStmt>
      </titleStmt>
      <editionStmt>
        <p n="1 Edizione">MURATORI, Antiquitates Italicae, VI, coll. 359-362;</p>
        <p n="2 Edizione">LAMI, I, pp. 321-2; BRUNETTI, II, 1, n. XXI, pp. 257-8;</p>
        <p n="3 Edizione">PL, XCVII, n. XVII, coll. 1014-5</p>
        <p n="4 Edizione">PASQUI, n. 15</p>
        <p n="5 Edizione">DD Kar., n. 150</p>
        <p n="6 Edizione">ChLA, XXV, n. 797</p>
        <p><signatures>Canonica,n.6</signatures>
        <note>Originale [A], A.C.Ar</note></p>
      </editionStmt>
      <publicationStmt>
        <date when="0783-08-09">783, ottobre 9</date>
        <textLang mainLang="la">latin</textLang>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <listBibl>
        <msDesc>
          <msIdentifier>
            <country>Germania</country>
            <region>Renania-Palatinato</region>
            <settlement type="town">Worms</settlement>
            <repository>Le carte della canonica di Arezzo</repository>
            <idno>15</idno>
          </msIdentifier>
          <physDesc>
            <summary>Carlo Magno, su richiesta di Ariberto, vescovo di Arezzo,
            conferma alla chiesa di San Donato tutti i suoi beni e in particolare il monastero di San Benedetto di
            Arezzo, fondato dal vescovo Cunimondo.</summary>
            <extent>
              <material>Pergamena rettangolare di taglio abbastanza regolare.Scrittura
              disposta secondo il lato lungo della pergamena. Inchiostro di colore grigio-bruno.</material>
              <dimensions unit="cm">
                <height>49,5 (48,5)</height>
                <width>31 (28,5)</width>
              </dimensions>
              <note>Pergamenadi colore bianco-giallastra, perfettamente conservata.</
note>
              <origDate notBefore="0835-10-06" notAfter="0843-08-29">853,ottobre
6-843,agosto 29</origDate>
            </extent>
          </physDesc>
          <typeDesc>
            <typeNote>A tergo, di mano del secolo IX: « Praeceptum Caroli de
monasterio Sancti Benedicti in civitate ».</typeNote>
          </typeDesc>
        </msDesc>
      </listBibl>
      <div1 type="ms">
        <div2 type="Protocollo" subtype="Invocatio">
          <p>(C)</p>
        </div2>
      </div1>
    </body>
  </text>
</teiHeader>

```

```

    </div2>
    <div2 type="Protocollo" subtype="Intitulatio">
      <p>
        <seg><rs type="person"><persName>Carolus</persName></rs> gratia Dei
<roleDesc>rex</roleDesc> Francorum et Langobardorum ac patricius Romanorum.</seg>
      </p>
    </div2>

    <div2 type="Protocollo" subtype="Inscriptio">
      <p>
        <seg></seg>
      </p>
    </div2>

    <div2 type="Protocollo" subtype="Salutatio">
    </div2>

    <placeName>
<date when="0783-08-09">783, ottobre 9</date>

      <settlement type="Town">Worms</settlement>
      <region></region>
    </placeName>
    <div2 type="Protocollo" subtype="Apprecatio">
    </div2>
<div3 type="Testo" subtype="Arenga">
  <p>
    Si petitionibus sacerdotum vel servorum Dei <lb />
    in quo nostris fuerint auribus prolatae libenter obaudimus et eas in Dei
nomine et effectui mancipamus, regiam consuetudinem exercimus et hoc nobis ad mercedem vel
stabilitatem<lb />
      regni nostri pertinere confidimus.
  </p>
</div3>

  <div3 type="Testo" subtype="Notificatio">
    <p>
      <seg>Igitur notum sit omnium fidelium nostrorum magnitudini presencium
videlicet et futurorum,</seg>
    </p>
  </div3>
  <div3 type="Testo" subtype="Inscriptio">
    <p>
      <seg>qualiter venerabilis vir<rs type="person"><persName>Aribertus</persName></rs>
<rs>, sanctae Arretianensis ecclesie <roleDesc>episcopus</roleDesc></seg>
    </p>
  </div3>
  <div3 type="Testo" subtype="Promulgatio">
  </div3>
  <div3 type="Testo" subtype="Narratio">
    <p>
      quae est constructa in honore sancti Donati, ad nostram accessit clemenciam et
petiit serenitati nostre ut omnes res ipsius ecclesie, quae ibidem a longo tempore et usque nunc visae
sunt per<lb>
      pertenuisse</lb>tenuisse, tam monasteria vel senodochia quamque ecclesias
baptismales seu reliquas possessiones quicquid <ref type="note" target="#a">a</ref> per donationes vel
confirmationes regum sive vinditiones, commutationes, quae ex bonorum<lb />
      hominum largitate ibidem datae vel condonatae sunt, vel etiam illud <gi
type="edificio">monasterium Sancti Benedicti</gi> infra ipsa civitate, quod bonae memoriae <rs
type="person"><persName>Cunemundus</persName></rs>, quondam <roleDesc>episcopus</roleDesc> antecessor
suus, legibus conparavit<lb />
      vel suo fundavit opere, una cum rebus suis propriis, quas de iure parentum
<gap reason="erase" unit="lettere" quantity="2"><ref type="note" target="#b">b</ref></gap> suorum
habuit et ad ipsam casam Dei visus fuit delegasse,
    </p>
  </div3>
  <div3 type="Testo" subtype="Dispositio">
    <p>
      <seg> inspectas ipsas preceptiones vel confirmationes sive vinditiones vel
commutationes<lb />
      denuo per nostram auctoritatem circa ipsum sanctum locum cedere et

```

confirmare deberemus. Cuius petitionem pro reverencia ipsius sancti loci nolimus denegare, sed in aelimosina nostra ita prestatissime et confirmasse cognoscite.<lb />

Praecipientes ergo iubemus ut, sicut constat ipsa casa Sancti Donati de predictis rebus a longo tempore iuste et rationabiliter legibus vestita fuisse et presenti tempore memoratus venerabilis vir <rs type="person"><persName>Aribertus</persName></rs><roleDesc>episcopus</roleDesc> recto trami<lb>

tramite</lb>te possidere videtur, inspectas ipsas praeceptiones vel confirmationes regum, ut diximus, seu vinditiones vel conmutationes sive traditiones bonorum hominum, ita in antea per hanc nostram auctoritatem<lb />

atque confirmationem valeat iam fatus <rs type="person"><persName>Aribertus</persName></rs> <roleDesc>episcopus</roleDesc> suique successores qui fuerint rectores ipsius sancti loci ipsis, ut supra memoravimus, rebus quieto ordine tenere et possidere et nullus quislibet de fidelibus <lb />

nostris amodo et deinceps sepe dicto episcopo vel successoribus suis de iam fatis rebus inquietare aut calomniam generare quoque tempore non presumatur, sed nostris et futuris temporibus ad ipsam casam Dei proficiant<lb /> in augmentis.</seg>

</p>

</div3>

<div3 type="Testo" subtype="Corroboratio">

<p>

<seg>Et ut hec auctorita firmior habeatur vel per tempora melius conservetur, <hi type="">[ma]nu</hi> propria subter eam decrevimus roborare et de anulo nostro [iussimus] sigillare.</seg>

</p>

</div3>

<div3 type="Testo" subtype="Formule">

</div3>

<div4 type="Escatocollo" subtype="Subscriptio Regis/Papae">

<p>

<seg>Signum (MF)<rs type="person"><persName>Caroli</persName></rs> gloriosissimi <roleDesc>regis</roleDesc>.</seg>

</p>

</div4>

<div4 type="Escatocollo" subtype="Recognitio">

<p>

<seg>Ercambaldus advicem Radoni ☩ (SR) <ref type="note" target="#c">c</ref> (SI)</seg>

</p>

</div4>

<div4 type="Escatocollo" subtype="Sottoscrizioni / Signa manuum">

<originDate>

<p> Da(ta) <num>VII</num> idus octobris, anno <num>XVI</num> et decimo regni nostri.</p></originDate>

<placeName>

Actum<settlement type="Town">Vurmasia</settlement>civitate.

<region>Renania-Palatinato</region>

<country>Germania</country>

</placeName>

</div4>

<div4 type="Escatocollo" subtype="Recognitio">

<p>In Dei nomi[ne] feliciter</p>

</div4>

<div type="notes">

<note xml:id="a">a quicquid: così</note>

<note xml:id="b">b seguono 2 lettere erase</note>

<note xml:id="c">c segue in note tironiane Ercambaldus advicem Radoni recognovi et subscripsi. Fulradus ambasciavit</note>

</div>

</div1>

</body>

</text>

</TEI>

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI SYSTEM "tei_ms.dtd">
<TEI
  xmlns="http://www.tei-c.org/ns/1.0">

  <teiHeader>

    <fileDesc>

      <titleStmt>
        <title></title>
        <funder> </funder>
      </titleStmt>

      <msDesc>
        <msIdentifier>
          <altIdentifier>
            <idno xml:id="numero">26</idno>
          </altIdentifier>

          <placeName>
            <settlement type="Town">Arezzo</settlement>
            <region>Tuscany</region>
          </placeName>
          <repository>Le carte della canonica di Arezzo</repository>

          <msContents>
            <textLang mainLang="la">latin</textLang>
          </msContents>

          <signatures>Canonica,n.13</signatures><note>F a l s o  in  forma  di  o r i g i n a l e  [F],
A.C.Ar.</note>

          <publisher>Marta Calleri</publisher>

          <editionStmt>
            <p n="1 Edizione">Dragoni,pp18-19(alla data 840 circa)</p>
            <p n="2 Edizione">Pasqui, n. 30 (alla data 840?)</p>
            <p n="3 Edizione">Calleri, pp.100-101</p>
          </editionStmt>
        </msIdentifier>
        <physDesc>
          <typeDesc>
            <summary>Pietro I, vescovo di Arezzo, costituisce la canonica dotandola di alcune terre.</
summary>
            <typeNote>A tergo, di mano del secolo XI: « Privilegium Petri episcopi de aliquantis
terris ».</typeNote>
          <extent>
            <dimensions unit="cm">
              <height>55</height>
              <width>48</width>
            </dimensions>
          </extent>
          <condition> Rettangolare di taglio abbastanza regolare. Rigato. Scrittura disposta secondo il
lato lungo della pergamena. Pergamena di colore biancastro, ben conservata, salvo alcuni forellini al
centro delle rr.6 e 13 e alla fine della r.10; inchiostro di colore bruno, inchiostro rossiccio nelle
sottoscrizioni.</condition>
        </typeDesc>
      </physDesc>
    </msDesc>
  </fileDesc>
</teiHeader>
<text>

  <body>
    <div1 type="Protocollo" subtype="Invocatio">

```

```

    <p>
      <seg>☐☐</seg>
    </p>
  </div1>
  <div1 type="Protocollo" subtype="Intitulatio">
    <p>
      IDEO EGO<rs type="person">Petrus</persname>,SERVUS SERVORUM DOMINI, SANCTE ARITINE AECCLIESIAE
      HUMILIS EXIGUUS EPISCOPUS.
    </p>
  </div1>
  <div1 type="Protocollo" subtype="Promulgatio">
    <p>
      MANIFESTUS SUMNO
    </p>
  <div1 type="Protocollo" subtype="Inscriptio">
    <p></p>
  </div1>

  <div1 type="Protocollo" subtype="Salutatio">
  </div1>
  <origDate notBefore="835, ottobre 6" notAfter="843, agosto 29">853,ottobre 6-843,agosto 29</
  origDate>
    <placeName>
      <settlement type="Town">Arezzo</settlement>
      <region>Tuscany</region>
    </placeName>
  <div1 type="Protocollo" subtype="Apprecatio">
  </div1>

  <div2 type="Testo" subtype="Arenga">
  </div2>
  <div2 type="Testo" subtype="Promulgatio">
  </div2>
  <div2 type="Testo" subtype="Narratio">
    <p>
      QUIA DOMNO ET PERPETUO <rs type="person">LOTHARIO</persname> SEMPER AUGUSTO<lb/>
      sepe nos admonuit cum cunctis obtimatibus atque doctores et sapientibus in eiusque imperio
      consistentibus, ut unusquisque episcopos <ref type="note" target="#a">a</ref> su☐ aeccliesi☐ canonica
      constituere, sicuti in Dei nomine ipsius piis<lb>
      piissimi<lb>simi augusti obaudientes precepta, nostr☐ aeccliesi☐ canonica construximus.
    </p>
  </div2>
  <div2 type="Testo" subtype="Dispositio">
    <p>
      Pro quibus utillimum prospeximus, una cum cunctos <ref type="note" target="#b">b</ref>
      sacerdotio, clero, cuncteque aeccliesi☐ nostr☐ volun<lb>
      voluntate</lb>tate et consensu, confirmare previdimus ad eadem canonica de rebus
      episcopatu nostro, quatenus ipsi canonici et sacerdotes, qui ibidem <ref type="note" target="#c">c</
      ref> consistunt et Domino famulantur, absque indigen<lb>
      indigentia</lb>tia vivere possint et delectabiliter exorare valeant, pro stabilitate regni
      et imperii dominorum nostrorum atque sancta aeccliesia et cuncto populo christiano eorumdem <ref
      type="note" target="#d">d</ref> a Domino regendum concessum <ref type="note" target="#a">e</ref>,<lb/>
      et ut eius semper adaugeatur plebs et populo. Primum quidem dedimus atque confirmavimus
      ibidem vineam unam qui dicitur ad Pinum, que est iuxta ipsam canonicam, seu campo suptus<lb/>
      ipsam vineam, qui similiter vocatur <ref type="note" target="#f">f</ref> ad Pinum, quidem
      et <ref type="note" target="#g">g</ref> prato qui dicitur et vocatur ad Salam <ref type="note"
      target="#h">h</ref>, idcirco et curticellam unam qui vocatur ad Clane, cum omnia ad se pertinentes,
      quidem et aliam curti<lb>
      curticellam</lb>cellam qui vocatur ad Moionam, de quantum Ursus actor ad suam tenet manus ad
      regendum, quamobrem et aeccliesiam Sancti Petri in Castello <ref type="note" target="#i">i</ref>
      quem <ref type="note" target="#j">j</ref> domnus Lotharius augustus pro suam mercedem <ref type="note"
      target="#k">k</ref><lb/>
      per suum confirmavit preceptum beati Donati marteris <ref type="note" target="#l">l</ref>.
      Iste vero predictae curticelle cum omnis <ref type="note" target="#m">m</ref> hominibus ibidem
      aspicientibus, servi pro servi, liberi pro liberi, concessimus, in eadem vide<lb>
      videlicet</lb>licet rationem dominandum sicuti usque actenus dominati sunt et
      pertenu☐runt <ref type="note" target="#n">n</ref>. De vero personis hominum dedimus ibidem personas
      quattuor manuales, pistores quoque et cocum atque lav[an]<lb>
      lav[an]darium</lb>darium et bifulcum.Haec omnia <ref type="note" target="#o">o</ref> ibidem
      perpetualiter usque ad voluntate domni et seniori nostro <ref type="note" target="#p">p</ref>
      imperatorii dedimus ad abendum in alimonia <ref type="note" target="#q">q</ref> et usu iam sepredictis
      canonicis, ut neque<lb>
      a nobis Petrus, indignus episcopus, neque ad subsequentes pontifices nullu modo auferatur
    
```



```

neque substraatur, sed semper in eorum potestate ad stipendia consistant diurnis et perpetuis tem<lb>
temporibus</lb>poribus, sicut cum cuncto sacerdotio, clero omnique aecclesiꝫ nostrꝫ propria
voluntate, commune consensu confirmavimus, ut olim dixissimus usque ad voluntate domni imperatorii <lb>
>
    et eius confirmatione.
</p>
</div2>
<div2 type="Testo" subtype="Formule">
</div2>
<div3 type="Escatocollo" subtype="Subscriptio Regis/Papae">
</div3>
<div3 type="Escatocollo" subtype="Recognitio">
</div3>
<div3 type="Escatocollo" subtype="Sottoscrizioni / Signa manuum">
<p></p>
<listWit>
    <wittnes>Ego<rs type="person"><persname>Petrus</persname>, humilis<roleDesc>episcopus</
roleDesc>, manu mea subscripsi.</wittnes>
    <wittnes id="Consenzienti">Ego<rs type="person"><persname>Gumfrid</
persname><roleDesc>archipresbiter</roleDesc> manu mea suscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Gumprand</persname><roleDesc>archidiaconus</
roleDesc> manu mea suscrisi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Pettrus</persname><roleDesc>diaconus</roleDesc>
manu mea subscripsi.</wittnes>
    <wittnes><ref type="note" target="#r">r</ref> Ego<rs type="person"><persname>Godelpetrus</
persname><roleDesc>diaconus</roleDesc> manu mea subscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Lampertus</persname><roleDesc>diaconus</roleDesc>
manus <ref type="note" target="#s">s</ref> mea subscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Iohanni <ref type="note" target="#t">t</ref> </
persname><roleDesc>presbiter</roleDesc> manu mea subscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Lupus <ref type="note" target="#u">u</ref>
diaconus manu me<a> subscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Prando</persname><roleDesc>presbiter</roleDesc>
manu mea subbscrisi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Gumpertu</persname><roleDesc> presbiter</roleDesc>
manu mea suscripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Uschisi</persname><roleDesc> presbiter</roleDesc>
manu mea scripsi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Raimualdu</persname><roleDesc> presbiter</
roleDesc> manu mea suscrisi.</wittnes>
    <wittnes>Ego<rs type="person"><persname>Barricinus</persname><roleDesc> presbiter</
roleDesc> manu mea scripsi.</wittnes>
</listWit>
<origDate notBefore="835, ottobre 6" notAfter="843, agosto 29">853,ottobre 6-843, agosto 29</
origDate>
<placeName>
    <settlement type="Town">Arezzo</settlement>
    <region>Tuscany</region>
</placeName>
</div3>
<div type="notes">
<note xml:id="a">a: episcopus: così</note>
<note xml:id="b">b: cunctos: così; con s aggiunta in seguito</note>
<note xml:id="c">c: ibidem: ibi corretto su rasura</note>
<note xml:id="d">d: eorumdem: o in soprallinea </note>
<note xml:id="e">e: segue, a capo, eraso est</note>
<note xml:id="f">f: vocatur: con r corretta su s</note>
<note xml:id="g">g: et: corretto su rasura</note>
<note xml:id="h">h: Salam: S aggiunta in seguito</note>
<note xml:id="i">i: Castello: stel corretto su rasura</note>
<note xml:id="j">j: quem: con segno abbreviativo superfluo sulla e</note>
<note xml:id="k">k: pro suam mercedem: così</note>
<note xml:id="l">l: marteris: così</note>
<note xml:id="m">m: omnis: così</note>
<note xml:id="n">n: pertenuꝫrunt: runt su rasura; nt in nesso</note>
<note xml:id="o">o: segue eraso iure</note>
<note xml:id="p">p: nostro: così</note>
<note xml:id="q">q: alimonia: la prima a corretta su e</note>

```

```
<note xml:id="r">r: il signum crucis è preceduto da ☩;Ego erasi </note>
<note xml:id="s">s. nostro: così</note>
<note xml:id="t">t: Iohanni: così</note>
<note xml:id="u">u: Lupus: la prima u corretta su a a cui segue m erasa.</note>
</body>
</text>
</TEI>
```

### **3 – Materiale introduttivo a XML-TEI da includere nella manualistica**

#### SEZIONE MANUALISTICA & DOWNLOAD

il materiale seguente è propedeutico alla comprensione del progetto TEI (text encoding initiative) per la codifica dei testi e contiene una breve descrizione, ad uso didattico ed interno, delle parti fondamentali di un documento XML che rispetti lo standard TEI.



<?xml?>



# TEI - Text Encoding Initiative

## Text Encoding Initiative

Un' introduzione allo standard per la riproduzione e la trasmissione dei testi in formato digitale





<?xml?>



## Perché TEI ?

- Lo schema TEI è il risultato finale degli sforzi ventennali di definizione di uno standard internazionale che garantisce la validità scientifica della codifica, l'intercambiabilità e la portabilità dei testi in formato elettronico.
- Progetto avviato nel 1987 dalle tre maggiori associazioni mondiali di studiosi di scienze umane attraverso metodologie informatiche, la *Association for Computers and the Humanities* (ACH) la *Association for Computational Linguistics* (ACL) e la *Association for Literary and Linguistic Computing* (ALLC)



<?xml?>



## Caratteristiche di TEI

- Anche se il progetto nasce nel 1989, la prima DTD è stata presentata nella pubblicazione uscita nel 1994 col titolo Guidelines for Electronic Text Encoding and Interchange (TEI P3).
- Per la sua universalità, la DTD di TEI comprende oltre 500 elementi (xml elements) e 100 attributi XML, la cui documentazione stampata > 2000 pag.
- TEI è giunta alla versione P5, più una versione LITE e versione specifiche, ad esempio TEI MS for manuscripts description
- Sito di riferimento: <http://www.tei-c.org>



<?xml?>



## Il sottoinsieme TEI LITE

- esso dovrebbe includere la maggior parte dell'insieme 'fondamentale' di marcatori della TEI, dal momento che questo contiene elementi rilevanti virtualmente per tutti i testi e per tutti i tipi di elaborazione testuale;
- esso dovrebbe essere in grado di trattare adeguatamente una varietà di testi ragionevolmente ampia, al livello di dettaglio incontrato nella pratica già esistente;
- esso dovrebbe essere utile sia per l'elaborazione di nuovi documenti sia per la codifica di quelli già esistenti;
- esso dovrebbe essere utilizzabile con un ampio spettro di applicazioni XML già esistenti;
- esso dovrebbe essere derivabile dall'intera DTD TEI, usando il meccanismo di estensione descritto nelle Norme;
- esso dovrebbe essere tanto conciso e semplice quanto consentito dalla conformità agli altri obiettivi.



<?xml?>



# Modularità di TEI

- Le Guidelines identificano tre gruppi di Tag sets:
  - **Core Tag set**: contiene il nucleo minimo di elementi comuni ad ogni tipo di testo, questo Tag set deve essere necessariamente incluso in ogni DTD.
  - **Base Tag set**: contiene diversi Tag set specifici per vari tipi di documenti fondamentali: prosa, poesia, dramma, parlato, lessicografico, terminologico. Questi tag set vanno inclusi nella DTD a seconda della tipologia dell'opera da codificare.
  - **Additional DTD Tag set**: insiemi di elementi per la rappresentazione di caratteristiche utili ad evidenziare determinate peculiarità del testo a partire da prospettive analitiche o applicazioni specializzate quali: codifica di apparati di varianti, rappresentazione di strutture ipertestuali, codifica di fonti primarie ed altro.



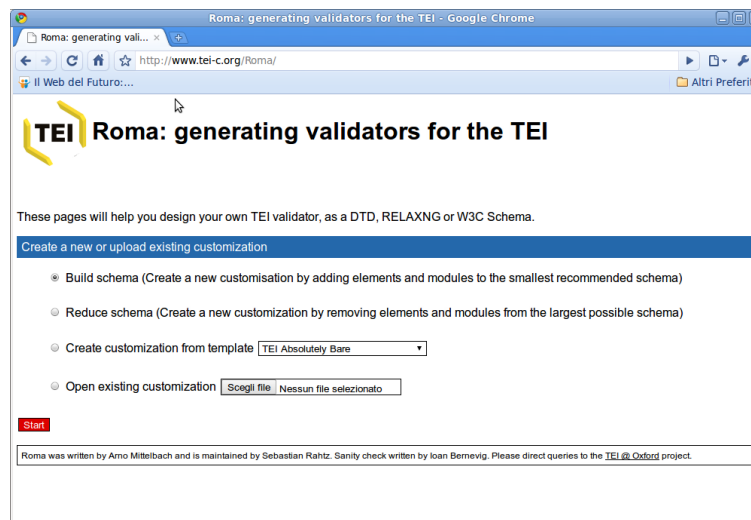


<?xml?>



# Roma: strumento di costruzione

- Ogni utente può costruire una sua versione della DTD di TEI:
- *Core tag set* + *base tag set* (a scelta) + *additional tag set* (uno o più)
- Roma ( <http://www.tei-c.org/Roma/> ) è lo strumento online per costruire il proprio set di marcatori TEI:





<?xml?>



# Struttura base di un documento TEI

■ ogni testo codificato conformemente alle specifiche della TEI è costituito da due parti:

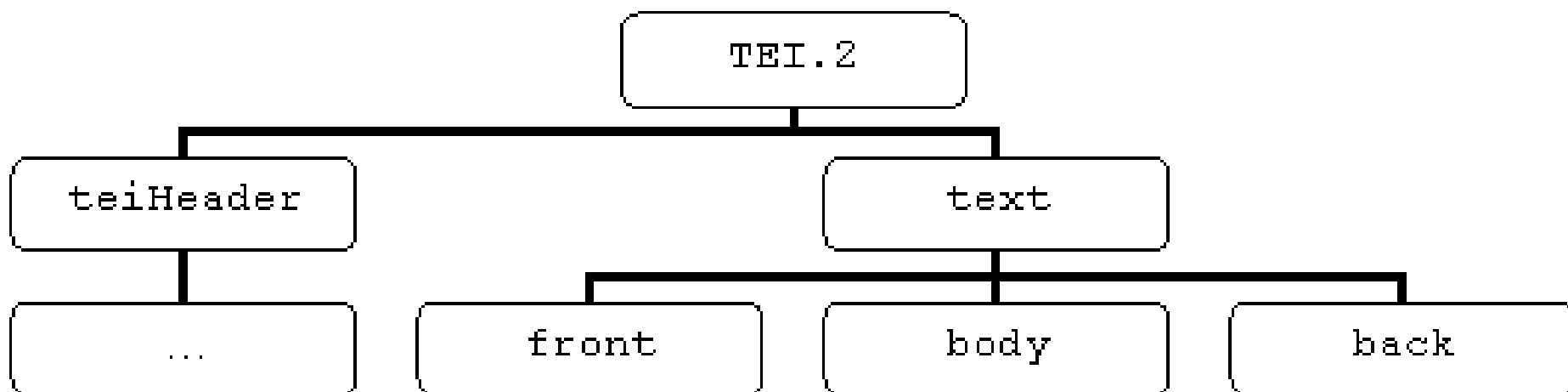
- un **TEI header** (codificato con l'elemento <teiHeader>), contenente le informazioni editoriali concernenti il documento elettronico,
- e un **TEI text** (codificato con l'elemento <text>) contenente la trascrizione codificata del testo in versione integrale.
- Ogni TEI header consta di **quattro parti**, una delle quali deve necessariamente ricorrere in ogni testo codificato secondo le raccomandazioni della TEI, ed è l'elemento <fileDesc> contenente i METADATI, mentre le altre tre, <encodingDesc>, <profileDesc>, <revisionDesc>, sono opzionali
- L'elemento text si divide, a sua volta, in quattro elementi:
  - <front> (opzionale), <body> (obbligatorio), <group> (obbligatorio) e <back> (opzionale).



<?xml?>



# Schema della struttura TEI





<?xml?>



## Lo schema minimo TEI è dunque:

<TEI.2>	[start tag della Tei]	
<teiHeader>	[informazioni della testata TEI]	
</teiHeader>		
<text>	[start tag del testo]	
<front>	[materiali dell'avantesto, ovvero dati che precedono il corpo del documento]	</front>
<body>	[corpo del testo]	</body>
<back>	[materiali annessi, dati che seguono il corpo del documento]	</back>
</text>	[end tag del testo]	
<TEI.2>	[end tag della Tei]	



<?xml?>



# Gli elementi dello schema

- **<front>** contiene qualsiasi elemento (testate, frontespizio, prefazioni, dediche, etc.) che si trovi prima dell'inizio del testo vero e proprio.
- **<group>** raggruppa un insieme di testi unitari o gruppi di testi.
- **<body>** contiene l'intero corpo di un singolo testo unitario, con l'esclusione di ogni elemento preliminare o di appendice.
- **<back>** contiene qualsiasi annesso o appendice che segua la parte principale di un testo



<?xml?>



# Struttura di un testo composito

- Anche un testo composito può essere dotato opzionalmente di avantesto e annessi. Nel mezzo si trovano uno o più gruppi di testi, ognuno eventualmente dotato di propri avantesti ed annessi.

```
<TEI.2>
  <teiHeader> [testata del testo composito] </teiHeader>
  <text>
    <front> [avantesto del testo composito] </front>
    <group>
      <text>
        <front>[avantesto del primo testo] </front>
        <body> [primo testo unitario] </body>
        <back> [annessi del primo testo] </back>
      </text>
      <text>
        <front>[ avantesto del secondo testo] </front>
        <body> [secondo testo unitario] </body>
        <back> [annessi del secondo testo] </back>
      </text>
      [altri testi o gruppi di testi]
    </group>
    <back> [annessi del testo composito] </back>
  </text>
</TEI.2>
```

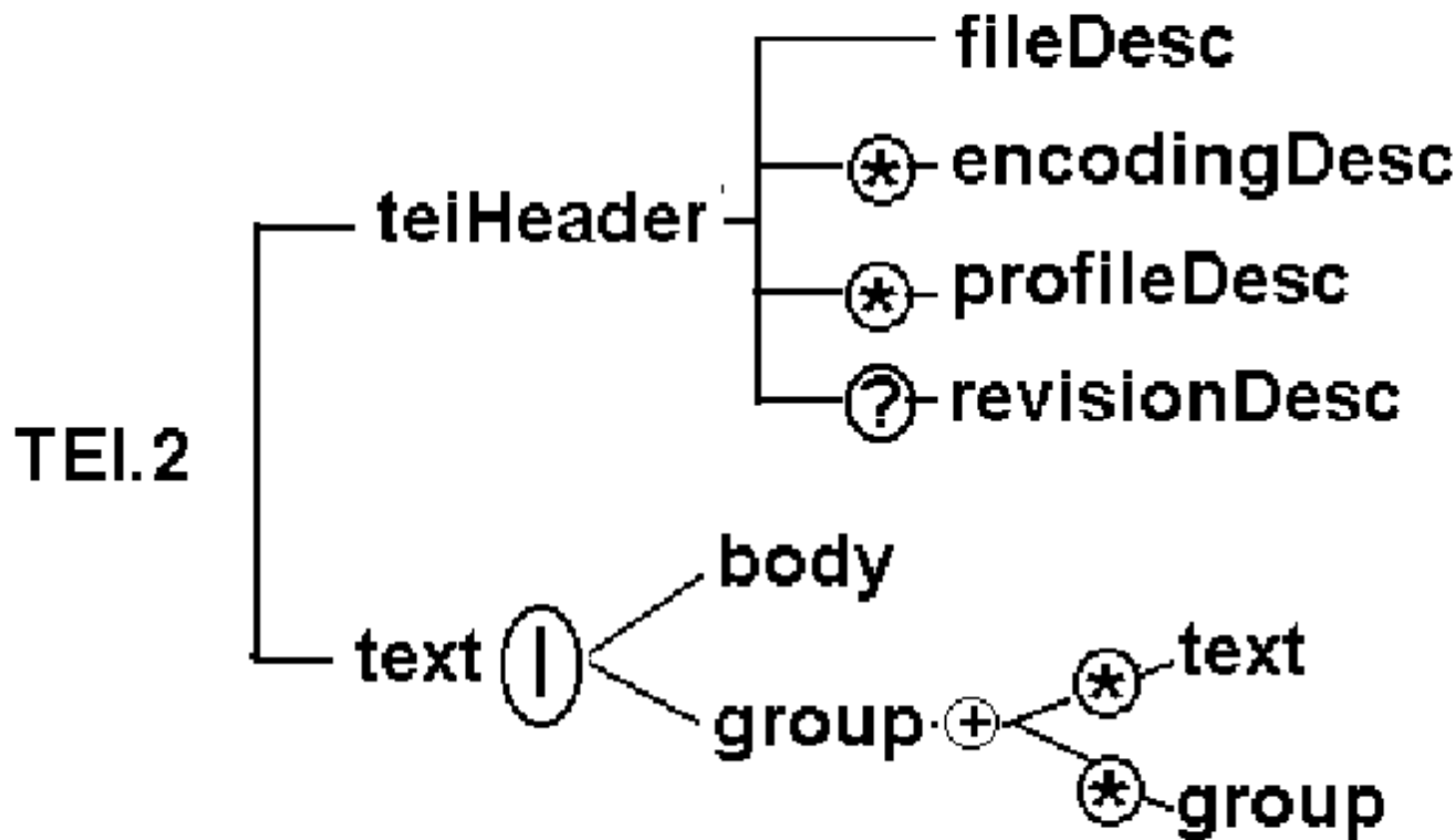




<?xml?>



# Schema TEI



A Group element must contain at least one text or group element



<?xml?>



## L'elemento <BODY>

- Il corpo di un testo in prosa può essere costituito solamente da una serie di paragrafi, oppure tali paragrafi possono essere raggruppati insieme in capitoli, sezioni, sottosezioni, etc.
- Nel primo caso, ogni paragrafo è marcato usando il marcatore <p>. Nel secondo caso, l'elemento <body> può essere diviso sia in una serie di elementi <div1>, sia in una serie di elementi <div>, i quali possono essere ulteriormente suddivisi





<?xml?>



# Marcatori di segmentazione

- `<p>` marca i paragrafi in prosa.
- `<div>` contiene una sezione dell'avantesto, del corpo o degli annessi di un testo.
- `<div1>` contiene una sezione di primo livello dell'avantesto, del corpo ed degli annessi di un testo (la più ampia, se `<div0>` non è usato, la seconda in ordine di grandezza altrimenti)
  - Qualora siano necessarie suddivisioni strutturali più piccole di una `<div1>`, quest'ultima può essere divisa in elementi `<div2>`, una `<div2>` in più piccoli elementi `<div3>`, etc., fino al livello `<div7>`. Se sono presenti più di 7 livelli della divisione strutturale, è necessario o modificare la DTD TEI per accettare `<div8>`, etc, oppure usare l'elemento non numerato `<div>`: un elemento `<div>` può essere suddiviso in elementi più piccoli `<div>`, senza limiti alla profondità di annidamento.

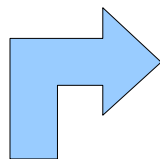


<?xml?>



# DTD TEI: elementi, attributi e entità

## ■ Un esempio:



<!ELEMENT anthology (poem+)>

<!ELEMENT poem (title?, stanza+)>

<!ELEMENT title (#PCDATA) >

<!ELEMENT stanza (line+) >

<!ELEMENT line (#PCDATA)>

```
<anthology>
  <poem><title>The SICK ROSE</title>
  <stanza>
    <line>O Rose thou art sick.</line>
    <line>The invisible worm,</line>
    <line>That flies in the night</line>
    <line>In the howling storm:</line>
  </stanza>
  <stanza>
    <line>Has found out thy bed</line>
    <line>Of crimson joy:</line>
    <line>And his dark secret love</line>
    <line>Does thy life destroy.</line>
  </stanza>
</poem>
  <!-- more poems go here -->
</anthology>
```



<?xml?>



# Contenuti e indicatori di occorrenza

- `<!ELEMENT title (#PCDATA) >`
- Parsed character data = l'elemento può contenere qualsiasi tipo di caratteri validi ma non altri elementi
- `<!ELEMENT poem (title?, stanza+)>`
- Indicatori di occorrenza:
  - + = almeno un'occorrenza dell'elemento
  - \* = 0, 1 o più volte
  - ? = 0 o 1, non di più



<?xml?>



# Connettori e gruppi di modelli

<!ELEMENT poem (title?, stanza+)>

<!ELEMENT poem (title? | stanza+)>

- La virgola rappresenta la sequenza
- La barra verticale | rappresenta l'alternanza
- Elementi multipli e la loro sequenza:

<!ELEMENT couplet (firstLine, secondLine) >

- E combinare tutto insieme:

<!ELEMENT poem (title?, (stanza+ | couplet+ | line+) ) >

<!ELEMENT poem (title?, (stanza | couplet | line)+ ) >

<!ELEMENT line (#PCDATA | name)\* >



<?xml?>



# Specificare gli attributi

- Se nel documento trovo:

```
<poem id='P1' status="draft"> ... </poem>
```

- Dovrò dichiarare gli attributi nella DTD:

```
<!ATTLIST poem
```

```
    id      ID          #IMPLIED
```

```
    status (draft | revised | published) "draft"
```

```
>
```



<?xml?>



# Valori di attributi

<!ATTLIST poem id ID #IMPLIED

status (draft | revised | published) #REQUIRED >

- CDATA = il valore dell'attributo contiene caratteri
- ID = il valore dell'attributo deve essere una singola stringa alfanumerica che inizia con un carattere alfabetico ed è identificatore unico
- **Valore di default:**
- #REQUIRED = obbligatorio specificarlo
- #IMPLIED = non è obbligatorio specificare il valore

<poem id='P1' status="draft"> ... </poem>



<?xml?>



## Dichiarazioni di entità

<!ENTITY tei "Text Encoding Initiative">

- Richiamabile nel documento con &tei;
- Entità interne ed esterne (system e public)

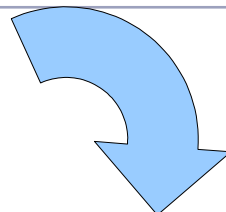
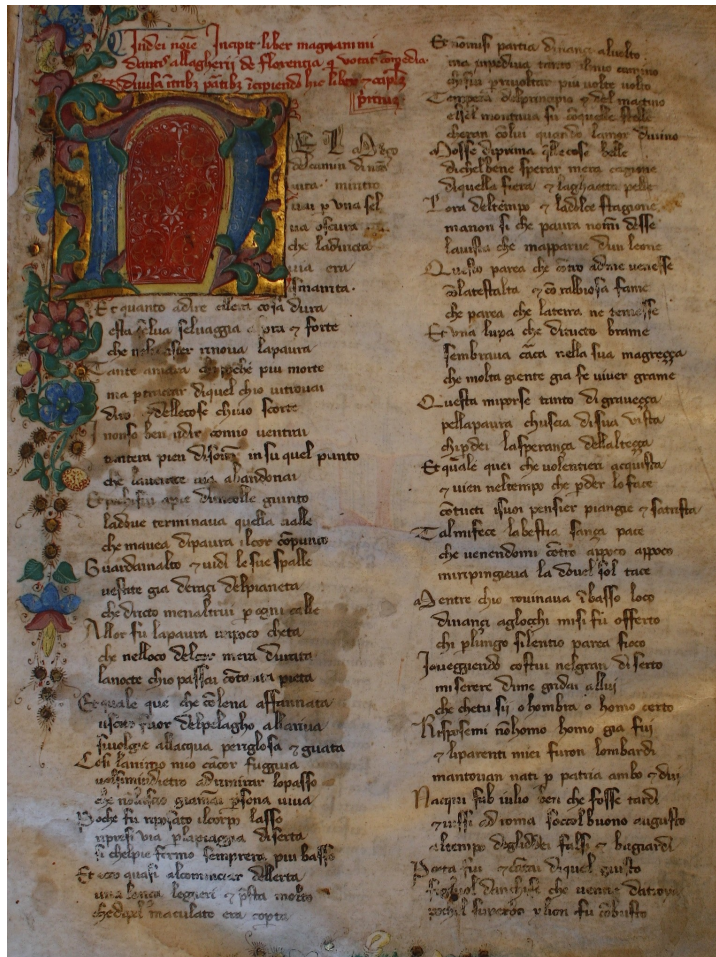
<!ENTITY ChapTwo SYSTEM "p4chap2.xml">

<!ENTITY p3.sg PUBLIC "-//TEI//TEXT Guidelines  
Chapter on XML//EN" "p4chap2.xml">

- Caratteri accentati: “è” come entità sarà:

<!ENTITY eacute "&#x00E9;">

<!-- LATIN SMALL LETTER E WITH ACUTE -->



## TEI Encoding - XML

```
<text>
  <body lang="ita">
    <div1 id="Inferno" type="parte">
      <head>INFERNO</head>
      <div2 id="Inferno.I" type="canto">
        <head>I</head>
        <lg type="terzina">
          <l>Nel mezzo del cammin di nostra vita</l>
          <l>mi ritrovai per una selva oscura,</l>
          <l n="3">ché la diritta via era smarrita.</l>
        </lg>
        <lg type="terzina">
          <l>Ahi quanto a dir qual era è cosa dura</l>
          <l>esta selva selvaggia e aspra e forte</l>
          <l n="6">che nel pensier rinova la paura!</l>
        </lg>
        <lg type="terzina">
          <l>Tant'è amara che poco è più morte;</l>
          <l>ma per trattar del ben ch'i' vi trovai,</l>
          <l n="9">dirò de l'altre cose ch'i' v'ho scorte.</l>
        </lg>
      </div2>
    </div1>
  </body>
</text>
```

Cortona (AR), Biblioteca dell'Accademia Etrusca





<?xml?>



## URL utili all'interno del sito TEI:

- Il sito [www.tei-c.org](http://www.tei-c.org) contiene materiale aggiornato relativo a TEI P5. In particolare segnaliamo:
  - **TEI: P5 Guidelines:** <http://www.tei-c.org/Guidelines/P5/>
  - **Roma:** Strumento online per produrre DTD TEI personalizzate: <http://www.tei-c.org/Roma/>
  - **TEI@SOURCEFORGE:** repository di tutti i files TEI aggiornati ed elenco versioni : <http://tei.sourceforge.net/>
  - **XSL stylesheets for TEI XML:** <http://www.tei-c.org/Tools/Stylesheets/>
  - **TEI WIKI:** <http://wiki.tei-c.org/index.php/Category:Tools>



<?xml?>



## Breve bibliografia:

- Questa presentazione costituisce un'introduzione alle caratteristiche della Text Encoding Initiative ed è stata elaborata esclusivamente come materiale didattico per il corso di “Linguaggi di Marcatura dei Testi” della Facoltà di Lettere e Filosofia di Arezzo (Univ. Siena). Poco di originale è stato aggiunto rispetto a quanto già presente in:
  - Griselda Online, i linguaggi di codifica:  
[http://www.griseldaonline.it/informatica/manuale\\_parte6\\_1.htm](http://www.griseldaonline.it/informatica/manuale_parte6_1.htm)
  - La DTD TeiLite, introduzione alla codifica dei testi Lou Burnard - C. M. Sperberg-McQueen Traduzione italiana di: Fabio Ciotti, Guendalina Demontis, Giuseppe Gigliozzi, Massimo Guerrieri, Andrea Loreti Revisione e cura traduzione italiana di: Fabio Ciotti Gennaio 1998 Revisione versione XML a cura di: Francesca Tomasi -  
[http://www.griseldaonline.it/informatica/tutorial\\_nuovi/TEI-XML\\_nuovo.pdf](http://www.griseldaonline.it/informatica/tutorial_nuovi/TEI-XML_nuovo.pdf)
  - Ciotti F. 2005, *Il manuale TEI LITE*, Edizioni Sylvestre Bonnard
  - Gli esempi utilizzati per illustrare il funzionamento della DTD sono tratti da:  
<http://www.tei-c.org/release/doc/tei-p4-doc/html/SG.html>