

DANILO ROMEI

AUTOGRAFI DIGITALI

Banca Dati “Nuovo Rinascimento”  
[www.nuovorinascimento.org](http://www.nuovorinascimento.org)

---

immesso in rete il 27 febbraio 2009

Il titolo più appropriato di questa comunicazione\* sarebbe *Autografi digitalizzati*. Di autografi propriamente “digitali” – cioè digitali in origine, se la cosa è possibile – conosciamo soltanto quelli compresi nel progetto *Digital Variants*, che fa capo all’Università di Edimburgo e a Domenico Fiormonte, all’indirizzo:

<http://www.selc.ed.ac.uk/italian/digitalvariants/home.htm>

e che tratta varianti redazionali di testi di autori viventi che scrivono al *computer*. Non è il nostro caso.

Nell’area di cui intendiamo occuparci opera la cosiddetta filologia “digitale” (o “computazionale”), che, oltre alle finalità tradizionali della disciplina, mira generalmente a creare *machine readable texts*.

Il significato di questa espressione deve essere radicalmente distinto da quello degli ordinari “testi elettronici” (*e-texts*), prodotti dai comuni applicativi di videoscrittura testuale (o ipertestuale). Questi applicativi generano *files* “binari” o *files* ASCII, redatti in base a un codice “proprietario” (come quello di Microsoft Word) o di “pubblico dominio” (come quello degli ipertesti HTML). In nessun caso l’elaboratore “capisce” il testo che sta processando.

Prendiamo, per esempio, un programma come Word. È un programma dalle enormi potenzialità, che l’utente comune spesso in gran parte ignora. Può controllare l’ortografia e la morfologia del testo che sta processando; può suggerire scelte lessicali; può effettuare la sillabazione delle parole; può creare indici e sommari automatici; può distinguere fra diversi livelli di struttura ecc. ecc. Si può dire allora che Word è un programma “intelligente”? Crediamo proprio di no, dal momento che non consente all’elaboratore di “capire” il testo. Si limita ad applicare degli algoritmi: che si tratti di fisica quantica o di poesia trobadorica il meccanismo è lo stesso. Inesorabile e stupido.

Invece sarebbe utile che l’elaboratore “capisce”, perché quanto più “capisce”

\* Si tratta appunto del testo di una comunicazione tenuta al Convegno di Studi *Edizione di autografi* in margine al Dottorato di ricerca in “Storia e tradizione dei testi nel Medioevo e nel Rinascimento” (Università di Firenze, 25-27 novembre 2004). Il testo – purtroppo – è invecchiato per alcune specifiche tecniche, ma la sostanza del ragionamento mi sembra ancora valida. Va da sé che il testo è stato concepito per una pubblicazione cartacea.

tanto più si rivela uno strumento efficace.

Perché “capisca” bisogna “insegnargli”. Naturalmente cominciando dalle nozioni elementari. Così, per esempio, bisogna dirgli che il testo che sta processando si chiama *Divina Commedia*, che si tratta di un *poema*, che ci sono tre *cantiche*, composte da *canti*, a loro volta divisi in *terzine* e in *versi*, che i versi sono *endecasillabi*, che *-ita* è una *rima*, che *Virgilio* è un *nome di persona*, che *Fiorenza* non è un *nome di persona* ma un *nome geografico* (la cui forma “normalizzata” è *Firenze*) e così via. Per questa via si crea un *machine readable text*.

A questo punto il *computer* è diventato “intelligente”? Forse quanto un lombrico (anzi, sicuramente meno). Però comincia a “balbettare”, cioè a fornire informazioni (“scalabili” – per così dire – a piacere) che non sono riducibili al riconoscimento di semplici stringhe di testo, come avviene per i *databases* testuali e per le concordanze elettroniche (che sono programmi “stupidi”), ma che appartengono ormai a un sistema di categorie logiche.

Naturalmente si può fare molto di più. E infatti negli scorsi decenni iniziative internazionali hanno elaborato strumenti molto più complessi, adattabili alle diverse branche del sapere.

In campo letterario si è affermata la codifica TEI (*Text Encoding Initiative*) (indirizzo del sito ufficiale del TEI Consortium: <http://www.tei-c.org>), uno *standard* internazionale che garantisce il pubblico dominio del codice, la sua uniformità, la sua durata (di gran lunga superiore a quella del *software* commerciale, anche se sempre da rapportare all’inevitabile obsolescenza dei prodotti informatici).

La codifica TEI è nata in SGML (*Standard Generalized Markup Language*), un linguaggio di programmazione elaborato dalla International Organization for Standardization (URL: <http://www.iso.org>) e pubblicato nel 1986 (è identificato dalla sigla ISO: 8879). L’SGML è il linguaggio internazionale della scienza; è un linguaggio potentissimo, complicatissimo, rigorosissimo, astratto. È anche quasi privo di *tools*, di strumenti di sviluppo e di utilizzo accessibili agli utenti di base. Per queste ragioni la TEI si è rivolta in un secondo tempo all’XML (*eXtensible Markup Language*), un *subset* di SGML, che si sta affermando in modo massiccio, specie per l’impiego in rete, e che resta comunque compatibile con il linguaggio di origine.

Noi, tuttavia, per la documentazione che intendiamo presentare, non adottiamo la codifica TEI bensì la codifica EAD (*Encoded Archival Description*), nata dalla collaborazione tra la US Library of Congress e la Society of American Archivists (il sito ufficiale è all’indirizzo <http://lcweb.loc.gov/ead/>). Anche la codifica EAD nasce in SGML e migra verso XML. È conforme allo *standard* ISAD(G) (*General International Standard Archival Description*), adottato dalla Ad Hoc Commission on Descriptive Standards (Stockholm 1993); è conforme allo *standard* ISO (e Unicode); è compatibile con la codifica TEI e con lo *standard* MARC delle biblioteche americane; è compatibile con il protocollo Z39.50, il formato di scambio fra le biblioteche di tutto il mondo.

La sua funzione non è quella di produrre inventari (di solito realizzati in forma di *databases*), bensì *finding aids* ipertestuali, cioè quelli che gli archivisti italiani chiamano “mezzi di corredo”, finalizzati al recupero di informazioni (*data retrieval*)

che di solito gli inventari non forniscono.

La scelta della codifica EAD è dettata dalla natura documentaria dei testi che adduciamo ad esempio: quattro lettere autografe di Pietro Aretino a Giovanni de' Medici (il celebre condottiero noto come Giovanni delle Bande Nere), scritte da Reggio – a nostro parere – nel 1524, conservate nel fondo Mediceo Avanti il Principato dell'Archivio Storico di Firenze (ASF MAP VI 97, cc. 808-809; CXXI 415; CXXII 106, cc. 106-106[bis]; CXXII 296, cc. 299-299[bis]). L'intero Mediceo Avanti il Principato è ora disponibile *on line* in veste di immagine digitale nel sito *internet* dell'Archivio di Stato di Firenze all'indirizzo:

<http://www.archiviodistato.firenze.it>

Per accedervi è necessario registrarsi. Le lettere – come si vede dalle segnature – non costituiscono una serie, ma sono composte da fogli sciolti, dispersi in filze diverse. Della storia del testo non si sa nulla. Si può dire soltanto che in passato facevano parte di unità archivistiche diverse da quelle in cui si trovano adesso, come si evince dalle concorrenti numerazioni delle carte.

Tutt'e quattro sono state edite da noi in forma elettronica nella Banca Dati Telematica “Nuovo Rinascimento” all'indirizzo:

<http://www.nuovorinascimento.org>

in cinque formati diversi, tutti funzionali – se non indispensabili – nell'ottica di una pubblicazione elettronica *on line*. Per la documentazione completa, come per le informazioni di contorno, rinviamo senz'altro ai *files* di “Nuovo Rinascimento”. In questa sede ci limitiamo a illustrare il testo più breve (MAP CXXI 415) con tutte le (odiose) limitazioni inerenti a una riproduzione cartacea.

L'interfaccia della pubblicazione *on line* è in formato HTML (*HyperText Markup Language*), il più comune formato di quella parte di *internet* che si suol chiamare *World Wide Web*. Dalla *home page* si accede a una premessa metodologica, all'indice e ai documenti di contorno con le informazioni essenziali.

Del testo si fornisce anzitutto un'immagine digitale (per la quale rimandiamo senz'altro al MAP *on line*).

Si fornisce quindi una trascrizione che si potrebbe dire “facsimilare” e che si sforza di riprodurre fin nei minimi dettagli la veste dell'originale, per quanto lo consentono gli strumenti disponibili. La sua funzione è di supporto alla lettura diretta dell'immagine digitale. Il formato prescelto è il PDF, il formato dei documenti generati da Adobe Acrobat (leggibili con Acrobat Reader, fino alla *release* 5.5, o con Adobe Reader per le *releases* superiori), che consentono un notevole livello di “formattazione” tipografica e garantiscono una visualizzazione pressoché uniforme su qualsiasi *client* (di qualsiasi piattaforma) siano riprodotti. Qui si riporta con qualche inevitabile perdita di qualità (fra parentesi tonde si sciolgono i compendi):

**Lettera di Pietro Aretino a Giovanni de' Medici  
da Reggio, [1524]**

ASF: MAP.CXXI.415

[mezzo foglio; missiva originale autografa: reca traccia evidente della piegatura originaria (in tre e poi di nuovo in tre, con la faccia esterna più logora) e del bollo di cerallacca]

[recto]

Jll.<sup>mo</sup> S.<sup>re</sup> p(er) commadamento [*sic*] d(i) V. S. lotto mando ammantoua | p(er) i denarj  
· i quali ha riceuti et aquella mandati <sup>gli haria</sup> segli persona | hauessi atta aportargli  
trouata. Jo / p(er) far bene / no ho uo | luto ch(e) gli mandi no(n) sapendo p(er) chj.  
poi no(n) so come, p(er) lauia | si uadi. V Jll. S. sidegni comandare quello ch(e) dessi  
si faccia

Di nouo la s<sup>ra</sup> Paula . a VS . si recomanda. et a xxij hore | carca di sospiri è ita aletto  
. et ha tenuta il moretto sino a | iij hore di notte aragionar(e). o pensate quanto gli saria  
| caro lesser co(n) voi

la vo(stra) speranza et Dea no(n) ha uoluto ch(e) bart<sup>o</sup>. si parta | anzi incasa  
ella et lamadre la uoluto ritenere . co(n) dir ch(e) | po l(a) V . S . disporre  
inmaggior cosa et cazouinculo saluando l'honor(e) di m(esser) gia(n) Pietro

[fondo pagina]

Semp(re) S.<sup>uo</sup> P<sup>o</sup> Aretino

[verso]

[sovrascritta]

·/. Al vnico Sre Joh.nes | De Medicj . immortalissimo | Cap.no et mio Patron(e) |||| ·/.  
Jn la via.

Segue una trascrizione “normalizzata”, cioè marcatamente orientata verso l’ortofonia e l’ortografia, nella logica di una impaginazione familiare al lettore moderno. Questa trascrizione è in formato TXT (“solo testo”) ed è la base per i successivi sviluppi. In questo caso il formato è riproducibile:

Lettera di Pietro Aretino a Giovanni de' Medici  
da Reggio, [1524]

ASF: MAP.CXXI.415

[recto]

Illustrissimo Signore, per commandamento di Vostra Signoria Lotto mandò a Mantova per i denari, i quali ha riceuti ed a quella mandati gli aría s'egli persona avessi atta a portargli trovata. Io, per far bene, non ho voluto che gli mandi non sapendo per chi; poi non so come per la via si vadi. Vostra Illustrissima Signoria si degni comandare quello che d'essi si faccia.

Di novo la signora Paula a Vostra Signoria si raccomanda ed a 22 ore, carca di sospiri, è ita a letto ed ha tenuta il Moretto sino a 3 ore di notte a ragionare. O pensate quanto gli saría caro l'esser con voi.

La vostra speranza e dea non ha voluto che Bartolommeo si parta; anzi in casa ella e la madre l'ha voluto ritenere, con dir che pò la Vostra Signoria disporre in maggior cosa e cazzovinculo, salvando l'onore di messer Gian Pietro.

Sempre servo Pietro Aretino

[verso]

[sovrascritta]

A l'unico Signore Johannes de' Medici, immortalissimo Capitano e mio Patrone, in la via

Quindi la trascrizione TXT è stata tradotta in linguaggio HTML, che fornisce il testo destinato alla normale lettura in *internet*. La visualizzazione è identica; non così il codice (peraltro semplicissimo):

```
<html>
<head>
  <title>Lettera di Pietro Aretino a Giovanni de' Medici,
    da Reggio, [1524]</title>
</head>
<body bgcolor="White" link="black">
  <table align="center" border="1" cellpadding="50"
    cellspacing="0" width="600">
    <tr>
      <td>
        <br>&nbsp;<br>
        <hr>
        <hr size="5">
        <p align="justify">
          <font face="Arial" size="+1">
            Lettera di Pietro Aretino a Giovanni de' Medici
          <br>
            da Reggio, [1524]
          <br>
          <br>
          ASF: MAP.CXXI.415
```

```

</font>
</p>
<hr size="5">
<br>
<br>&nbsp;<br>
<br>&nbsp;<br>
<p align="justify">
  <font face="Arial">
    [recto]
    <br>&nbsp;<br>
    Illustrissimo Signore, per comandamento di Vostra
    Signoria Lotto mand&ograve; a Mantova per i denari,
    i quali ha riceuti ed a quella mandati gli
    ar&iacute;a s'egli persona avessi atta a portargli
    trovata. Io, per far bene, non ho voluto che gli
    mandi non sapendo per chi; poi non so come per la
    via si vadi. Vostra Illustrissima Signoria si degni
    comandare quello che d'essi si faccia.
    <br>
    Di novo la signora Paula a Vostra Signoria si
    recomanda ed a 22 ore, carica di sospiri, &egrave;
    ita a letto ed ha tenuta il Moretto sino a 3 ore di
    notte a ragionare. O pensate quanto gli
    sar&iacute;a caro l'esser con voi.
    <br>
    La vostra speranza e dea non ha voluto che
    Bartolommeo si parta; anzi in casa ella e la madre
    l'ha voluto ritenere, con dir che p&ograve; la
    Vostra Signoria disporre in maggior cosa e
    cazzovinculo, salvando l'onore di messer Gian
    Pietro.
    <br>&nbsp;<br>
    Sempre servo Pietro Aretino
    <br>&nbsp;<br>
    <br>&nbsp;<br>
    [verso]
    <br>
    [sovrascritta]
    <br>
    <br>
    A l'unico Signore Johannes de' Medici,
    immortalissimo Capitano e mio Patrone, in la via
  </font>
</p>
<br>&nbsp;<br>
<br>&nbsp;<br>
<table align="center" border="1" cellpadding="5">
  <tr>

```

```

        <td><center>&nbsp;<a href=" ../indice.html"><font
        face="Arial">indietro</font></a>&nbsp;</center>
        </td>
    </tr>
</table>
<br>&nbsp;<br>
</td>
</tr>
</font>
</table>
</body>
</html>

```

Prima di passare alla versione XML è necessario premettere qualche informazione elementare.

E cominciamo dall’SGML: un linguaggio di marcatura (*markup language*) nel quale il codice è interamente leggibile e non affidato a simboli astrusi, come avviene per il *software* commerciale. L’SGML genera infatti dei *files* ASCII all’interno dei quali i marcatori (*tags*) sono delle comuni espressioni verbali racchiuse fra parentesi angolari. Per esempio:

```
<title>Divina Commedia</title>
```

Una barra obliqua (*slash*) segnala la chiusura del marcatore: il testo marcato è quello compreso fra il *tag* e il relativo *tag* di chiusura.

L’SGML è un sistema di *generic markup*, ovverosia un linguaggio astratto che si occupa soltanto della definizione della struttura logica del documento, indicando gli elementi che la compongono e le relazioni che si instaurano al suo interno. La definizione è contenuta nella *Document Type Definition* (DTD). In realtà sarebbe più giusto dire che l’SGML è un metalinguaggio che fornisce le regole per costruire infiniti linguaggi, in grado di generare documenti universalmente “portabili”, cioè compatibili con tutti i sistemi informatici.

La definizione di un *set* specifico di marcatori (il “lessico” del linguaggio) è in questo caso riservata alla codifica EAD/XML.

L’XML, come si è detto, è un sotto-insieme dell’SGML, caratterizzato dalla estensibilità (può definire nuovi marcatori e nuovi attributi), dalla strutturazione gerarchica (con possibilità di “nidificazione”), dal rigore (l’XML esige che il documento sia “valido” [*valid*] e “ben formato” [*well-formed*], perché i *parsers* – ovvero le *utilities* preposte ad analizzare il documento da elaborare – rifiutano il codice “sporco”), dalla separazione tra la struttura dei dati (la sola trattata dall’XML) e la loro visualizzazione (delegata a un fratellino minore: l’XSL, così come i *links* ipertestuali, riservati all’XLL). L’XML si è affermato anche per la sua propensione a “fare da ponte” fra applicazioni e sistemi diversi.

Il documento XML si compone di tre parti:

1. la *XML Declaration*, che dichiara quale sia la “specificata” XML utilizzata e quale sia il *set* di caratteri usato nella DTD e nel testo codificato
2. la *Document Type Definition* (DTD), ovvero la definizione del tipo di documento a cui appartengono i dati che si intendono elaborare e del *set* di marcatori che vi si trovano; può essere memorizzata nello stesso *file* del testo (prima del testo stesso) o trovarsi in un *file* esterno (in questo caso il testo codificato deve contenere un riferimento alla DTD su cui si modella)
3. la *Document Instance*, ovvero il testo codificato, in cui compaiono insieme i dati e il *markup* applicato ai dati in conformità alla sintassi prevista.

La DTD definisce:

1. gli elementi (*elements*), identificati dai marcatori
2. gli attributi (*attributes*) che si possono applicare agli elementi
3. i valori (*values*) che gli attributi possono assumere.

La sintassi è la seguente:

```
<elemento attributo="valore">testo</elemento>
<figlio sesso="maschio">Piero</figlio>
```

Con questo metodo è possibile marcare testi utilizzando DTD già esistenti oppure sviluppare nuove DTD idonee a descrivere ogni tipo di documento. Nel caso specifico utilizziamo la DTD EAD 1.0, adattata al linguaggio XML seguendo le indicazioni di varie università e centri di ricerca americani, con particolare riguardo alla *EAD and EBIND Technical Information* degli University of Illinois Archives

(<http://web.library.uiuc.edu/ahx/ead/tech/default.asp>)

e alle *EAD Help Pages* dell’University of Virginia Web Site

(<http://jefferson.village.virginia.edu/ead/>).

Poiché si tratta di un archivio virtuale ci siamo concessi ulteriori adattamenti e semplificazioni. Ci siamo, peraltro, astenuti dal creare nuovi elementi.

Ed ecco la struttura concreta del documento XML. Ne commentiamo in modo sommario le componenti di maggior rilievo. Per facilitare la lettura del codice abbiamo rinunciato ai caratteri “estesi”, ovvero a quei caratteri che non si trovano nell’alfabeto inglese (come le lettere accentate) e che devono essere rappresentati mediante espressioni del tipo “&#x00E0;” (per il carattere à). Va da sé che nel codice “autentico” i caratteri “estesi” sono indispensabili.

Si comincia con la *XML Declaration*:

```
<?xml version="1.0" encoding="UTF-8" ?>
```

Si continua con la *Document Type Definition* (DTD):

```

<!DOCTYPE ead PUBLIC "-//Society of American
Archivists//DTD ead.dtd (Encoded Archival Description
(EAD) Version 1.0)//EN" "ead/ead.dtd" [

<!ENTITY % eadnotat PUBLIC "-//Society of American
Archivists//DTD eadnotat.ent (EAD Notation
Declaration)//EN" "ead/eadnotat.ent">
%eadnotat;

<!ENTITY % ISOLat1 PUBLIC "-//ISO 8879-1986//ENTITIES
Added Latin 1//EN" "char_ent/ISOLat1.ent">
%ISOLat1;

]>

```

nella quale si specifica che si sta utilizzando la DTD EAD [DOCTYPE ead] e che si tratta di una DTD di pubblico dominio [PUBLIC] elaborata dalla [Society of American Archivists], che si trova materialmente collocata in un *file* esterno di nome [ead.dtd]; non si rinvia, peraltro, al sito ufficiale della codifica EAD, nel quale si trova il modello (e come sarebbe più semplice fare), ma si rinvia a un *file* locale nella *subdirectory* [/ead] (per evitare il traffico di rete). Si specifica che la versione che stiamo applicando è la 1.0 [Version 1.0] e che la codifica è in inglese [EN].

La versione 1.0 della codifica EAD è ancora in SGML; per tradurla in XML dobbiamo apportare delle modifiche che sono definite nel *file* [eadnotat.ent]; dobbiamo anche adottare un *set* di caratteri idoneo alla lingua italiana, *set* definito nel *file* [ISOLat1.ent].

La *document instance* è racchiusa fra i marcatori <ead>.....</ead>. A questo punto si avvia un processo di “nidificazione”, ovvero di inclusione di elementi dentro altri elementi, in base a una struttura gerarchica la cui sintassi è rigorosamente definita dalla DTD (e dalle regole dell’XML).

Gli elementi di livello gerarchico immediatamente inferiore a <ead> (e in esso contenuti) sono <eadheader> e <archdesc>.

Il primo costituisce l’“intestazione” del documento e fornisce le informazioni relative al *file* precettate dalle *guidelines* della codifica EAD:

```

<eadheader findaidstatus="edited-full-draft"
langencoding="ISO 639-2" audience="internal" id="a0">
<eadid type="XML Record" />
<filedesc>
<titlestmt>
<titleproper extent="all">MAP.CXXI.415</titleproper>
<author>Finding aid prepared by Danilo Romei.
</author>
</titlestmt>

```

```

</filedesc>
<profiledesc>
  <creation>
    Finding aid encoded by Danilo Romei on <date>23
    novembre 2003</date>
  </creation>
  <language>
    <language>English</language>
  </language>
</profiledesc>
</eadheader>

```

Vi si dice, anzitutto, che lo stato di elaborazione del documento [findaidstatus] è completo (è sempre possibile apportare delle variazioni, che però ogni volta devono essere registrate), che la lingua della codifica è compatibile con lo *standard* [ISO 639-2], che la destinazione del documento [audience] non è pubblica. All'“intestazione” è assegnato un identificatore di livello 0 [id="a0"].

È necessaria una precisazione sull'attributo [audience] e sul suo valore [internal]. È chiaro che la situazione normale di un *finding aid* EAD è quella di un archivio reale, in cui sono programmabili documenti destinati al pubblico e documenti riservati al personale (“interni”, appunto). Nel caso di un archivio digitale *on line* – com'è quello di cui si ragiona – il visitatore si arresta all'interfaccia HTML; i documenti XML sono predisposti per l'elaborazione dei dati da parte del *server* (in questo senso si tratta di una destinazione “interna”); il visitatore ha accesso ai dati (mediante indici, *forms*, *reports* ecc. in HTML), ma non ai documenti. E infatti nel nostro caso non è stato predisposto nessun sistema di traduzione visuale dei documenti XML; se si aprono con un *browser* appare il codice nudo.

Nell'“intestazione” si aggiunge il titolo del *file* [MAP.CXXI.415], il creatore del *file* (noi medesimi) e la data di creazione [23 novembre 2003].

Con l'elemento <archdesc> (dedicato alla “descrizione archivistica”) si entra nel vivo. Al suo interno due elementi di pari livello: <did> e <add>.

Con il primo si forniscono le informazioni relative all'unità archivistica (quella “reale”, il foglio sciolto con la lettera dell'Aretino conservato nell'Archivio di Stato di Firenze) prescritte dallo *standard* ISAD(G): deposito, ente conservatore, soggetto produttore, titolo, data di creazione, descrizione fisica, *abstract*, ecc. Considerata la funzione esemplificativa del documento, si è operata qualche semplificazione.

L'elemento <add> dovrebbe introdurre informazioni marginali non previste dallo schema ISAD(G). Al contrario – in maniera del tutto eterodossa, ma non incompatibile con il modello di codifica e senza innovare alcun elemento – si è collocata qui, come <relatedmaterial>, la trascrizione normalizzata del testo, sottoponendo anch'essa alla codifica EAD. In questo modo si è resa possibile l'elaborazione meccanica del documento, processabile come rudimentale *machine readable text*.

Si sono applicati tre soli marcatori: <persname> (nome di persona),

<geogname> (nome geografico), <date> (data). Per tutt'e tre è previsto l'attributo [normal] che introduce come valore la forma normalizzata dell'espressione a testo. Per esempio, il nome *Johannes de' Medici*, che compare nella sovrascritta, viene così codificato:

```
<persname normal="Medici, Giovanni de'">Johannes de'
Medici</persname>
```

in questo modo si rende processabile la forma normalizzata [Medici, Giovanni de'] e si rende processabile la variante *Johannes de' Medici* sotto la forma normalizzata.

Si sono inserite anche due annotazioni (marcatore <note>) e due riferimenti bibliografici (marcatore <bibref>, con la giunta dell'elemento interno <edition>).

Ed ecco il codice del documento nella sua interezza:

```
<?xml version="1.0" encoding="UTF-8" ?>

<!DOCTYPE ead PUBLIC "-//Society of American
Archivists//DTD ead.dtd (Encoded Archival Description
(EAD) Version 1.0)//EN" "ead/ead.dtd" [

<!ENTITY % eadnotat PUBLIC "-//Society of American
Archivists//DTD eadnotat.ent (EAD Notation
Declaration)//EN" "ead/eadnotat.ent">
%eadnotat;

<!ENTITY % ISolat1 PUBLIC "-//ISO 8879-1986//ENTITIES
Added Latin 1//EN" "char_ent/ISolat1.ent">
%ISolat1;

]>

<ead>
<eadheader findaidstatus="edited-full-draft"
langencoding="ISO 639-2" audience="internal" id="a0">
<eadid type="XML Record" />
<filedesc>
<titlestmt>
<titleproper extent="all">MAP.CXXI.415</titleproper>
<author>Finding aid prepared by Danilo Romei.
</author>
</titlestmt>
</filedesc>
<profiledesc>
<creation>
Finding aid encoded by Danilo Romei on <date>23
```

```

    novembre 2003</date>
  </creation>
  <language>
    <language>English</language>
  </language>
</profiledesc>
</eadheader>
<archdesc level="item" type="inventory" langmaterial=
"ita">
  <did id="a1">
    <head>Overview of the Record</head>
    <repository label="Repository:">
      <corpname>Archivio di Stato di Firenze</corpname>
    </repository>
    <origination label="Creator:">
      <persname normal="Aretino, Pietro">Pietro Aretino
      </persname>
    </origination>
    <unittitle label="Title:">Lettera di Pietro Aretino a
    Giovanni de' Medici, da Reggio</unittitle>
    <unitdate label="Dates:">1524</unitdate>
    <physdesc>
      <genreform>Un mezzo foglio: missiva originale
      autografa [reca traccia evidente della piegatura
      originaria (in tre e poi di nuovo in tre, con la
      faccia esterna piÙ usurata) e del bollo di
      ceralacca]</genreform>
    </physdesc>
    <abstract label="Abstract:">Commissioni finanziarie.
    Notizie dell'innamorata del destinatario.</abstract>
    <unitid label="Record Item:" countrycode="IT"
    repositorycode="ASF">MAP.CXXI.415</unitid>
  </did>
  <add>
    <relatedmaterial>
      <head>Testo</head>
      <p>
        Lettera di Pietro Aretino a Giovanni de' Medici
        <lb />
        [da Reggio]
        <note>
          <p>[La localizzazione geografica è ricavabile con
          certezza dai riferimenti circostanziali]</p>
        </note>
        <lb />
        [1524]
        <note>
          <p>[La data è ricavabile con certezza dai

```

riferimenti circostanziali|</p>  
 </note>  
 </p>  
 <p>  
 [recto]  
 <lb />  
 <lb />  
 <persname normal="Medici, Giovanni de'">  
 Illustrissimo Signore</persname>, per comandamento  
 di <persname normal="Medici, Giovanni de'">Vostra  
 Signoria</persname> <persname normal="Lotto">Lotto  
 </persname> mandò a <geogname normal="Mantova">  
 Mantova</geogname> per i denari, i quali ha  
 riceuti ed a quella mandati gli aría s'egli  
 persona avessi atta a portargli trovata. Io, per  
 far bene, non ho voluto che gli mandi non sapendo  
 per chi; poi non so come per la via si vadi.  
 <persname normal="Medici, Giovanni de'">Vostra  
 Illustrissima Signoria</persname> si degni  
 comandare quello che d'essi si faccia.  
 <lb />  
 Di novo la signora <persname normal="Paula">Paula  
 </persname> a <persname normal="Medici, Giovanni  
 de'">Vostra Signoria</persname> si raccomanda ed a  
 22 ore, carca di sospiri, è ita a letto ed ha  
 tenuta il <persname normal="Moretto">Moretto  
 </persname> sino a 3 ore di notte a ragionare. O  
 pensate quanto gli saría caro l'esser con voi.  
 <lb />  
 <persname normal="Paula">La vostra speranza e dea  
 </persname> non ha voluto che <persname normal=  
 "Bartolommeo">Bartolommeo</persname> si parta; anzi  
 in casa ella e <persname normal="madre [di Paula/  
 Pavola]">la madre</persname> l'ha voluto ritenere,  
 con dir che pò la <persname normal="Medici,  
 Giovanni de'">Vostra Signoria</persname> disporre  
 in maggior cosa e cazzovinculo, salvando l'onore di  
 messer <persname normal="Gian Pietro, messere">Gian  
 Pietro</persname>.  
 <lb />  
 <lb />  
 Sempre servo <persname normal="Aretino, Pietro">  
 Pietro Aretino</persname>  
 <lb />  
 <lb />  
 [verso]  
 <lb />  
 [sovrascritta]

```

<lb />
A l'unico Signore <persname normal="Medici,
Giovanni de' ">Johannes de' Medici</persname>,
immortalissimo Capitano e mio Patrone, in la via
</p>
<p>
<bibref inline="true">COPIE: 1 copia ms. nel fondo
Bongi dell'Archivio di Stato di Lucca, p. 7r-v
(LARIVAILLE 1989 23)</bibref>
</p>
<p>
<bibref inline="true">
EDIZIONI:
<lb />
<edition>LARIVAILLE 1989 23-24 (non ne conosce
l'autografo; cfr. p. 23, n. a: "In assenza
dell'autografo [...]")</edition>
</bibref>
</p>
<p>
<bibref inline="true">
BIBLIOGRAFIA:
<lb />
vedi: LARIVAILLE 1989 23-24
</bibref>
</p>
</relatedmaterial>
</add>
</archdesc>
</ead>

```

Orbene, che cosa abbiamo ottenuto?

È evidente che ci siamo fermati a un livello bassissimo di codifica: tre soli marcatori. Eppure, se si creasse un ipertesto EAD/XML che contenesse le tremila lettere dell'Aretino (e le loro varianti) così codificate, avremmo una banca dati d'importanza eccezionale. Basterebbe implementare un motore di ricerca (imprescindibile in qualsiasi ipertesto codificato), programmato per effettuare interrogazioni sia *full text* su semplici stringhe di testo (come fanno i più stupidi *databases* testuali), sia sulla base dei marcatori, dei loro attributi, dei loro valori (con l'ausilio di operatori logici), per ottenere risposte inimmaginabili sui testi di carta. E non sarebbe più necessario compilare indici, incipitari, glossari, tavole di concordanza ecc.: una semplice *query* potrebbe fornire tutto questo – e tanto altro ancora – in pochi istanti.

Ma niente ci obbliga a fermarci qui. Anzi, ci pare scontato che un serio lavoro di ricerca dovrebbe procedere molto più avanti. E niente ci obbliga a fermarci ai *tags* previsti dalla DTD EAD. La caratteristica saliente del linguaggio XML è proprio l'estensibilità. Ci possiamo creare tutti i marcatori, gli attributi, i valori che ci servono; basta rispettare la procedura prevista, cioè la loro esatta definizione in una nuova

DTD.

Mettiamo che il ricercatore sia un metricista. Le lettere dell'Aretino allegano spesso dei testi poetici (come quella segnata MAP VI 797 che fa parte del piccolo *corpus* in questione). Ebbene, niente impedisce di creare i *tags*:

```
<metrica incipit="Io ch'un secolo e un mezzo ho
buggerato">
<metrica forma="sonetto_caudato">
<sonetto schema="ABBA.ABBA.CDC.DCD.dEE">
<endecasillabo ritmo:="3_6_8_10">
<rima -ato="buggerato:visitato:peccato:innamorato">
```

e così via, *ad libitum*. Diventano banali, adesso, incipitari, tavole metriche, rimari, tavole di frequenza ecc. Quello che conta è che ogni fenomeno (metrico, in questo caso) che ci interessa è documentabile non in base a campionature, ma in base a una esemplificazione integrale, che l'elaboratore, se richiesto, ci porge con la massima sollecitudine e precisione. E gli strumenti si affinano strada facendo, in risposta a specifiche sollecitazioni del testo.

A questo punto bisogna frenare. Non vogliamo fare la figura dell'imbonitore delle meraviglie a buon mercato. È ovvio che ogni medaglia ha il suo rovescio. Il limite più grave delle codifiche è rappresentato dal loro "costo". Che non è soltanto un costo finanziario, di risorse, di strutture e di attrezzature, ma soprattutto un costo umano: di tempo, di lavoro, di competenze. Anche mettendo in conto l'automazione delle procedure, la marcatura affinata di un testo di proporzioni imponenti è un'impresa immane. È quasi scontato che il progettista-coordinatore di una ricerca di questa natura debba essere coadiuvato (oltre che da un tecnico informatico) da operatori abbastanza qualificati da capire le particolarità tecniche del lavoro da compiere e nello stesso tempo sacrificabili in operazioni che hanno molto di meccanico. Per giunta i tempi di realizzazione di un eventuale progetto devono essere ragionevolmente accorciati, se non si vuole incappare nella rapida obsolescenza della strumentazione informatica. Tutto questo comporta un'organizzazione della ricerca affatto inconcepibile nel contesto dell'università italiana.

E qui ci fermiamo.