Dalla scrittura medievale alle edizioni digitali

L'esperienza del Codice Diplomatico digitale della Lombardia Medievale

Ada Grossi

1. Introduzione

Il Codice Diplomatico digitale della Lombardia Medievale (CDLM) nasce nel 2000 presso l'Università di Pavia sotto il patrocinio dalla Regione Lombardia e con la collaborazione degli altri Atenei lombardi.[1] Il progetto è stato ideato e coordinato fin dall'inizio da Michele Ansani, che ha portato l'edizione delle fonti documentarie medievali – e quindi Diplomatica e diplomatisti – nel vivo della "arena digitale": in più di un'occasione, man mano che idee e realizzazioni concrete prendevano forma, egli ne ha dato conto attraverso riflessioni critiche e metodologiche volte a inquadrare il CDLM nel panorama più ampio delle problematiche legate al trattamento elettronico dei testi e ne ha tracciato e precisato le linee-guida. [2]

Sotto la guida dello stesso Ansani, una *équipe* di giovani ricercatori, di cui chi scrive fa parte, è impegnata nella realizzazione concreta del CDLM. L'oggetto sono le fonti documentarie medievali, il territorio la Lombardia (non rigidamente intesa come la Regione attuale, ma come Lombardia storica), i secoli quelli dall'VIII al XII: in questo territorio e in questo arco cronologico il Codice contempla l'acquisizione e l'edizione digitale della documentazione tramandata dagli archivi di istituzioni sia ecclesiastiche che civili, oltre che di famiglie.

Il criterio fondamentale è quello della ricostruzione degli antichi archivi mediante un'operazione di censimento che tiene conto di tutti gli elementi a disposizione, soprattutto degli strumenti di corredo antichi e moderni, e prevede l'individuazione di pezzi eventualmente dispersi. Per il resto, il CDLM rientra nella tipologia dei tradizionali codici diplomatici territoriali:

1 URL: http://cdlm.unipv.it.

2 Si vedano in particolare: Michele Ansani, Diplomatica (e diplomatisti) nell'arena digitale. In: Scrineum. Saggi e materiali on-line di scienze del documento e del libro medievali 1 (1999), http://scrineum.unipv.it/ansani.html; una versione ridotta del medesimo saggio è comparsa a stampa in Archivio Storico Italiano 158 (2000), n. 584 (disp. 2), pp. 349–379; Michele Ansani, Il Codice diplomatico digitale della Lombardia medievale: note di lavoro. In: Dino Puncuh (a cura di), Comuni e memoria storica. Alle origini del comune di Genova, Atti del convegno, Genova 24–26 settembre 2001, Genova 2002 (Atti della Società Ligure di Storia Patria NS 42/1 (2002), pp. 23–49; Michele Ansani, Medium-evo. Gli studi medievali e il mutamento digitale, Firenze, 21–22 giugno 2001 (http://www.storia.unifi.it/_PIM/Medium-Evo); Michele Ansani, Diplomatica e nuove tecnologie. La tradizione disciplinare fra innovazione e nemesi digitale. In: Scrineum. Saggi e materiali on-line di scienze del documento e del libro medievali 1 (2003), http://dobc.unipv.it/scrineum/rivista/ansani.html.

"evidentemente con correzioni d'impianto rispetto al modello tradizionale e che, sebbene risulteranno stemperate dall'architettura ipertestuale (come si sa, ogni architettura ipertestuale riconfigura e moltiplica, più o meno orientandoli e disorientandoli, i percorsi della lettura e della consultazione), mirano a rappresentare qualcosa di diverso da una semplice strategia tecnico-editoriale; sia perché è ormai solo la disponibilità sistematica di fonti vagliate ed edite criticamente a poter alimentare e potenziare la ricerca documentaria e medievistica in generale per i secoli anteriori perlomeno al XIII; sia perché questa sistematicità insiste e deve anzitutto insistere e concentrarsi su aree di tradizione documentaria e giuridica omogenea; sia perché sembrerebbero le stesse potenzialità della rete e dei linguaggi digitali a consigliare l'adozione di un simile obiettivo programmatico". [3]

L'obiettivo, appunto, è creare un sistema omogeneo di accesso sia statico che dinamico al *corpus* documentario raccolto e da raccogliere, edito e inedito. L'accesso statico permette di sfogliare le pagine di un grande libro virtuale strutturato in "parti" corrispondenti alle aree territoriali (a cui si accede tramite una mappa stilizzata) e in "capitoli" corrispondenti alle singole edizioni. L'accesso dinamico, reso possibile dal trattamento elettronico dei testi, consente non solo di navigare in essi e tra di essi, ma mette anche a disposizione dell'utenza una varia strumentazione per la ricerca storico-documentaria (per esempio indici, repertori di notai, cronologie, in futuro si auspica anche riproduzioni fotografiche dei documenti) e soprattutto un sistema avanzato di ricerca automatica sui testi, sviluppato dal Centro di ricerche informatiche per le discipline umanistiche-signum della Scuola Normale Superiore di Pisa.[4]

Il CDLM è ospitato nel sito di "Scrineum. Saggi e materiali on-line di scienze del documento e del libro medievali"[5], rivista digitale fondata dall'Università di Pavia e arricchitasi via via di nuove esperienze; essa è dotata di *links* che la collegano ad altri siti (citiamo per tutti "Reti Medievali")[6], dispone di una sempre più ricca biblioteca e ospita ormai numerosi contributi specialistici. Recentemente il CDLM è stato inoltre associato ad altri progetti e strumenti promossi dalla Regione Lombardia attraverso il portale di risorse storiche e archivistiche "Lombardia Storica".[7]

2. Ragioni e opportunità di un'edizione digitale

Per comprendere appieno la novità e l'utilità che crediamo rappresentata dal CDLM è necessario, innanzitutto, precisare che per quanto numerose siano le edizioni di documenti medievali lombardi, una parte molto cospicua del

3 Ansani, Codice, testo corrispondente alla nota 18.
4 Si rimanda a http:/www.signum.sns.it/analisi_testuale/settore_informatico/tresy/index.html.
5 http://scrineum.unipv.it.
6 http://www.retimedievali.it.
7 http://plain.unipv.it.

patrimonio documentario conservato è ancora inedita: il CDLM è attivamente impegnato su questo fronte, pur dovendo fare i conti con una certa scarsità di ricercatori specializzati e di risorse economiche.

Quanto poi alle fonti edite, è opportuno rilevare quanto complesso e vario sia il panorama delle edizioni critiche esistenti. A parte ovvie considerazioni relative ai diversi livelli di qualità scientifica del materiale a disposizione, pubblicato in un arco cronologico che va dalla stagione degli eruditi ai giorni nostri, va sottolineato che i criteri di compilazione delle opere edite sono molteplici e diversi. Ci sono le edizioni per fondi, cioè di singoli fondi archivistici: ma se alcune di esse partono dalla ricostruzione, per quanto possibile, degli antichi archivi, altre considerano solo i fondi di conservazione attuali (escludendo così tutto ciò che, pur originariamente conservato insieme, ha poi seguito per i motivi più vari strade archivistiche diverse); ci sono i cartari di istituzioni ecclesiastiche e monastiche specifiche che recuperano il materiale prodotto da un ente dovunque sia conservato; ci sono i codici diplomatici che raccolgono la documentazione riguardante un certo territorio in un determinato intervallo di tempo, come il "Codex diplomaticus Langobardiae" del Porro Lambertenghi[8]; ci sono edizioni che danno conto di una tipologia documentaria precisa, come i "Placiti del Regnum Italiae" di Cesare Manaresi[9]; oppure ancora opere come gli "Atti del Comune di Milano", iniziati dallo stesso Manaresi[10] per il XII secolo e la prima parte del XIII e poi proseguiti per tutto il Duecento da Maria Franca Baroni[11]: tale monumentale edizione, nel suo complesso, costituisce una sorta di codice diplomatico dell'istituzione comunale milanese (anche se in esso vengono considerati solo gli atti prodotti dal comune e non quelli da esso ricevuti o diversamente pertinenti).[12]

8 Giulio Porro Lambertenghi, Codex Diplomaticus Langobardiae (Historiae Patriae Monumenta 13), Torino 1873.

9 Cesare Manaresi, I placiti del Regnum Italiae (Fonti per la storia d'Italia 92, 96, 97), Roma 1955–1960.

10 Cesare Manaresi, Gli atti del comune di Milano fino all'anno *MCCXVI,* Milano 1919.

11 Maria Franca Baroni, Gli atti del comune di Milano nel secolo XIII, vol. 1: 1217–1250, Milano 1976; Maria Franca Baroni/Roberto Perelli Cippo, Gli atti del comune di Milano nel secolo XIII, vol. 2/1: 1251–1262, Alessandria 1982; Idem, Gli atti del comune di Milano nel secolo XIII, vol. 2/2: 1263–1276, Alessandria 1987; Maria Franca Baroni, Gli atti del comune di Milano nel secolo XIII, vol. 2/3: Indici volume 2 (1251–1276), fonti-bibliografia, Alessandria 1987; Eadem, Gli atti del comune di Milano nel secolo XIII, vol. 3: 1277–1300, Alessandria 1992; Eadem, Gli atti del comune di Milano nel secolo XIII, vol. 3/1: Appendice 1211–sec. XIII. Indici, fonti e bibliografia, Alessandria 1992; Eadem, Gli atti del comune di Milano nel secolo XIII, vol. 4/1: Appendice 1176–sec. XIII, Milano 1997; Eadem, Gli atti del comune di Milano nel secolo XIII, vol. 4/2: Indice dei nomi di persona e di luogo, elenco dei documenti editi nell'opera, Milano 1998.

12 È stato rilevato che i criteri di selezione adottati per queste edizioni pongono qualche difficoltà per lo studio tanto dell'istituzione quanto della procedura, come notava Thomas Behrmann nel 1995 (Thomas Behrmann, Von der Sentenz zur Akte. Beobachtungen zur Entwicklung des Prozeßschriftgutes in Mailand. In: Hagen Keller/Thomas Behrmann (Hg.), Kommunales Schriftgut in Oberitalien: Formen, Funktionen, Überlieferung, München 1995, pp. 71–90): a seguito di tali osservazioni fu poi pubblicato un volume specificamente dedicato a una particolare tipologia di documenti comunali nel XIII secolo, che esulano quindi comunque dall'ambito operativo del CDLM, cfr. Maria Franca Baroni, Gli atti di 'querimonia' tra i documenti giudiziari del Comune di Milano (sec. XIII), Alessandria 1997.

Questi brevi cenni sono sufficienti per comprendere la complessità e varietà del panorama editoriale: si tratta di opere disomogenee tra loro che è parso quindi fondamentale riunire in un *corpus* unico, insieme alle nuove edizioni che vengono progressivamente realizzate sia in seno a questo progetto che al di fuori di esso (naturalmente previo consenso degli autori).

In questi anni la possibilità di trattare elettronicamente i testi, in generale, ha mosso risorse e iniziative soprattutto da parte dei grandi archivi e delle grandi biblioteche: anche in Italia, seppure in ritardo rispetto ad altri paesi in Europa e in America. A riguardo si pensi ad alcune tra le istituzioni più prestigiose del settore, che hanno prodotto soprattutto sistemi di riconversione[13]: basti qui citare gli "elektronischen Monumenta Germaniae Historica" (eMGH), o, per restare in ambito strettamente lombardo, i due CD-rom che contengono i molti volumi dell'edizione degli "Atti del Comune di Milano" nei secoli XII e XIII che abbiamo citato sopra.[14]

Se questi sono esempi di edizioni digitalizzate, il CDLM è invece un'edizione digitale.[15] Esso non consiste nel semplice trasferimento su supporto informatico delle edizioni tradizionali (vecchie o nuove che siano) ma è invece un'elaborazione del materiale testuale in formato elettronico sulla base di uno *standard* di codifica. Come abbiamo già detto, le fonti non sono quindi disponibili semplicemente come testi statici con il solo vantaggio di essere raccolte tutte insieme e quindi di offrire una maggiore comodità di consultazione: la novità e la particolarità di un'edizione digitale risiede nella possibilità di compiere ricerche mirate (e pressoché illimitate) in quanto i testi sono stati preventivamente "marcati" mediante l'individuazione degli elementi di cui il testo si compone e le relazioni reciproche tra di essi. La differenza tra l'edizione digitalizzata e quella digitale, come la nostra, è molto rilevante. Per quanto i citati CD degli MGH o degli "Atti del Comune di Milano" siano molto utili e certamente molto comodi da consultare, essi non sono che un semplice trasloco di testi, un cambiamento del supporto da cartaceo a elettronico. Possiamo scegliere quale versione sia più conveniente utilizzare, ma

13 Per una sintetica rassegna delle realizzazioni più significative, cfr. Michele Ansani, Sull'edizione digitale di fonti documentarie. In: Roberto Greci (a cura di), Medioevo in rete tra ricerca e didattica (Itinerari medievali 5), Bologna 2002, pp. 38–39, e Gianmarco De Angelis, Repertorio critico di risorse digitali per gli studi di storia della scrittura latina e della produzione manoscritta nel Medioevo, http://dobc.unipv.it/scrineum/repertorio.

14 Cfr. il piano dell'opera http://www.mgh.de/emgh. Quanto a Gli atti del Comune di Milano nei secoli XII-XIII, Milano 2000, si tratta della conversione in formato PDF dei volumi di cui alla note 10 e 11: il trasferimento dei dati su supporto digitale è stato realizzato a cura dell'Università degli Studi di Milano, grazie alla collaborazione dell'Istituto di Storia del Diritto Italiano, del Dipartimento di Scienze della Storia e della Documentazione Storica e della Divisione Coordinamento delle Biblioteche.

15 Un caso "intermedio", destinato ad assumere caratteristiche più simili al CDLM, almeno in parte, può individuarsi per esempio negli "Anglo Saxon Charters", ai quali si può accedere *on line* attraverso una maschera di ricerca articolata: non è però ancora disponibile, anche se da tempo annunciato, il sistema di interrogazione della base dati mediante criteri diplomatistici (http://www.trin.cam.ac.uk/chartwww/NewRegReg.html).

i testi a disposizione sono esattamente gli stessi; né particolarmente innovativi sono gli strumenti di accesso ai testi medesimi: a parte, è ovvio, una maggiore rapidità e praticità di consultazione rispetto ai ponderosi volumi cartacei, il formato digitalizzato consente, al più, di procedere a una ricerca per stringa, per sequenza di caratteri (sistema peraltro che non sempre dà buoni risultati).[16]

In un'edizione digitalizzata, invece, data l'edizione di un testo, la struttura del testo medesimo viene evidenziata, cioè appunto "marcata", in modo tale da consentire di recuperare tutti gli elementi individuati. La ricerca per stringa non è che il più banale dei sistemi di ricerca: lo specifico meccanismo di *information retrieval* reso possibile da un motore di ricerca apposito, che opera sulla marcatura, consente ricerche mirate, potenzialmente illimitate e di natura molteplice.

Va da sé che "non si potrà ragionevolmente sostenere che un'edizione elettronica sia di per sé migliore di un'edizione a stampa: dovrà offrire qualcosa di più, qualcosa che l'edizione a stampa, un'edizione tradizionale non è in grado di offrire. E questo plus-valore non potrà certamente coincidere con una maggiore rapidità nelle procedure di pubblicazione"; al contrario, avrà bisogno "di una maggiore trasparenza nelle procedure critiche, di una maggiore consapevolezza e responsabilità dell'editore" e darà luogo alla "possibilità di contemperare molteplici livelli di accesso alla documentazione, calibrati tanto sulle esigenze della ricerca quanto sulle strutture documentarie – strutture che risultano storicamente definibili e definite". [17]

Il contenuto è responsabilità di chi codifica e la complessità della struttura richiede che siano degli specialisti a farlo: nel caso della documentazione inedita, poi, chi dà l'edizione e chi codifica sono la stessa persona.

Ponendosi il CDLM su tale piano, al momento della sua ideazione è stato necessario considerare preliminarmente le applicazioni *software* a disposizione, gli *standards* di trattamento elettronico dei testi già esistenti, nonché i linguaggi di codifica testuale: considerato che il CDLM si sarebbe impegnato a un livello estremamente specialistico, cioè di edizione delle fonti, quella di seguire orme altrui si è rivelata una strada poco praticabile ed è stato necessario elaborare soluzioni nuove e originali.

Sono stati dunque utilizzati *standards* disponibili e riconosciuti (il metalinguaggio XML che adotta una sintassi di tipo dichiarativo) ma si è rinunciato a modelli approntati da altri settori della comunità scientifica impegnati

16 Anzi, sia detto per inciso, il sistema della ricerca per stringa di testo a volte non funziona affatto, visto che qualunque carattere estraneo a quelli indicati nella sequenza da ricercare – segni di parentesi per scioglimenti, restituzioni o integrazioni, trattini di a capo etc. – interrompe la sequenza medesima e rende impossibile recuperare tutte le occorrenze: lanciando una ricerca per la stringa "Mediolanenses" non verranno individuate, per esempio, le occorrenze del tipo "Mediol(anenses)". Nel CDLM il problema è stato superato sostituendo il tradizionale utilizzo di segni di parentesi con l'impiego di colori diversi per la visualizzazione del testo compreso tra tali parentesi.

17 Ansani, Il Codice diplomatico digitale, testo corrispondente alla nota 11.

nella codifica di testi, dando invece vita a un modello del tutto autonomo. A più riprese lo stesso Ansani ha ribadito di avere individuato "nel *Vocabulaire International de la Diplomatique* lo strumento ideale per avviare una discussione e un confronto concreti all'interno della comunità dei diplomatisti, lo strumento da cui partire per fissare i requisiti di una *Document Type Definition* (DTD)", cioè un protocollo strutturato di riferimento nel quale sono fissate le regole di codifica, "da impiegare nell'edizione elettronica di testi documentari; che, in sostanza, sia per la comunità degli studiosi ed editori di fonti d'archivio medievali ciò che TEI è per filologi e letterati".[18]

I documenti medievali, di cui ci occupiamo, sono infatti testi di tipo altamente formalizzato: su di essi, quindi, il trattamento elettronico risulta particolarmente efficace. La struttura di codifica deriva nel nostro caso dalla natura stessa delle fonti, documenti strutturati di per sé e la cui analisi è solidamente definita e fissata nella manualistica.[19] Come è ovvio, che si tratti di un contratto di vendita o di un privilegio papale o imperiale, il testo è comunque costruito per sua natura su schemi formali che, pur nella loro varietà, saranno sempre riconducibili a una griglia: sarà quindi sempre possibile riconoscere ed evidenziare la struttura del testo che, per quanto elastica, è comunque composta da elementi ben definiti che hanno relazioni altrettanto ben definite tra loro. In altre parole, la Diplomatica stessa ci provvede di un insieme strutturato di *standards* descrittivi che individuano architettura del documento, descrittori, elementi e relazioni reciproche.

Il riferimento per la marcatura è la già citata DTD (*Document Type Definition*), che non è altro che l'elenco strutturato dei marcatori nel quale sono compresi e messi in relazione tra loro tutti gli elementi di cui abbiamo bisogno: a) quelli che individuano le partizioni formali del documento (protocollo, testo, escatocollo e relative sotto-partizioni); b) quelli che segnalano le funzioni personali giuridiche e documentarie (autore, destinatario, rogatario etc.); c) quelli relativi alla marcatura di nomi di persona, di luogo e di istituzioni; d) gli interventi editoriali che riguardano la visualizzazione di note, restituzioni, scioglimenti e così via. La DTD è sufficientemente rigida, perché costituita da elementi definiti, ma anche sufficientemente elastica, perché si adatta plasticamente all'oggetto da descrivere. Tramite la DTD è dunque possibile descrivere qualsiasi documento medievale: essa costituisce una struttura universale, un modello teorico generale, all'interno del quale si descrive di volta in volta la casistica specifica e particolare. Si intende che la DTD è aggiornabile

18 Ansani, Codice, testo corrispondente alla nota 15.

19 Tra la vasta bibliografia a riguardo citeremo qui solo un'opera recente di importanza fondamentale, destinata per più di una ragione ad assumere un ruolo di riferimento (come ampiamente illustrato in Ansani, Diplomatica, testo corrispondente alle note 4–7, cui si rimanda): cfr. Maria Milagros Cárcel Ortí (a cura di), Commission Internationale de Diplomatique. Comité International des Sciences Historiques, Vocabulaire International de la Diplomatique, València 1994.

in qualsiasi momento: nel corso degli anni sono state apportate modifiche e integrazioni che si sono rivelate necessarie con il progredire dell'esperienza di codifica e con l'insorgere di nuove necessità.

Il meta-linguaggio utilizzato per la marcatura o codifica è XML (*eXtensible Mark-up Language*), che utilizza una sintassi standardizzata di tipo cosiddetto "dichiarativo" e consente quindi di definire a piacere il "vocabolario" dei singoli "tags" o marcatori, definiti e denominati sulla base delle esigenze scientifiche e della struttura dei testi.

Quanto al *software* necessario, per le operazioni di codifica viene utilizzato un *editor* di XML, *Note Tab*[20]: allo scopo di rendere più veloce l'inserimento materiale dei "tags" e per evitare errori di battitura è stata predisposta una libreria dei marcatori utilizzati dal CDLM che consente, con un semplice doppio *click*, di racchiudere automaticamente una selezione di testo entro un "tag" di apertura e uno di chiusura.

Il testo così marcato viene controllato da *XML Spy*, un *software* di categoria *parser*, in grado cioè di rilevare gli errori di sintassi eventualmente commessi, sia in generale che in riferimento alla DTD specifica.

3. La logica della codifica

Entrando nel vivo della codifica o marcatura dei testi, essa viene effettuata tramite i "tags" che individuano ciascun elemento: il testo da marcare viene semplicemente compreso tra un "tag" di apertura e uno di chiusura. La codifica, che contiene le informazioni che non vengono visualizzate ma che vengono utilizzate dal motore di ricerca, è quindi costituita da tutto quanto compreso tra parentesi angolari, cioè tra il "tag" di apertura con i suoi eventuali attributi, p. es. <REST>testo restituito</REST> oppure <PERSONA nm="Petrus de Puteo">Petrus de Puteo</PERSONA>, e quello di chiusura. In particolare vengono marcati toponimi, antroponimi e istituzioni che vengono gestiti in modo da generare automaticamente indici, liste di frequenza e altri strumenti. Inoltre, gli elementi marcati possono essere "annidati" l'uno dentro l'altro all'infinito, purché nel rispetto della DTD, ove sono fissate gerarchie e relazioni reciproche tra i diversi "tags" e ove sono definiti gli "attributi" previsti all'interno di ciascun "tag".

Le caratteristiche fondamentali di questo procedimento risiedono nella possibilità di individuare (marcare, appunto) gli elementi del testo direttamente su di esso lasciandoli esattamente come e dove si trovano, segnalando la presenza di ciascun elemento mediante i "tags" e i loro eventuali "attributi" (alcuni servono alla normalizzazione e regolarizzazione dei nomi, altri hanno lo scopo di segnalare relazioni tra elementi – per es. relazioni di parentela tra

20 Ai nostri fini è sufficiente la versione *Light*, ma ne esiste anche una versione *Pro* dotata di *browser* e di alcune altre *features* supplementari.

persone –, di precisare usi cronologici, di stabilire identificazioni – di luoghi e di persone –).

I vantaggi di tale procedimento di trattamento dei dati sono evidenti e risultano tanto più chiari se ne confrontiamo le caratteristiche con quelle del più diffuso procedimento di classificazione in uso, ovverosia un comune *database* relazionale generato con *Access*. In un simile *database* disponiamo di campi (le colonne) e di *records* di inserimento dati (le righe): il testo deve essere quindi scomposto in elementi (per esempio i nomi di persona e di luogo) che vengono poi forzati all'interno di uno schema rigido e del tutto estraneo a quello del testo di partenza. Uno dei limiti insuperabili più ovvii è che sarà comunque impossibile "annidare" campi gli uni dentro gli altri e quindi tenere conto delle posizioni relative dei dati e delle loro funzioni: e per quanto numerose siano le operazioni di filtro e selezione operabili sui dati inseriti, esse saranno comunque legate alla rigida struttura imposta dall'esterno, che non potrà mai riprodurre la struttura del testo originario.

Codificare cioè marcare un testo, al contrario, significa rispettare la complessità e ricchezza delle informazioni che esso contiene individuando le parti di cui si compone senza apportarvi alcuna modifica, né tantomeno scomporlo o destrutturarlo. Per usare una similitudine molto semplice ma efficace, marcare un testo digitale con XML è come analizzare il testo di una pagina stampata usando pennarelli evidenziatori di colori diversi (ogni colore corrisponde a un tipo di dati) ed etichette adesive recanti varie annotazioni, ove gli evidenziatori sono i "tags" e le etichette adesive sono gli "attributi".

4. La procedura di codifica

La codifica di un testo può innanzitutto essere suddivisa in tre blocchi: le istruzioni del *software* che assegnano a ciascun *file* il percorso di foglio di stile e DTD di riferimento (non visualizzate); i dati relativi a data, titolo, tipologia, luogo di conservazione e segnatura del documento (in parte visualizzati e in parte no) e quelli relativi all'editore (visualizzati in coda all'edizione); l'edizione del documento e relativi apparati (che contiene a sua volta il testo da visualizzare e le meta-informazioni definite dalla marcatura).

L'ultima parte, cioè l'edizione vera e propria del documento medievale, si articola in quattro fasi di codifica del testo:

– Strutture formali, partizioni e sotto-partizioni del documento. I marcatori definiscono lo svolgimento del discorso documentario nelle sue articolazioni principali, secondo un'architettura di riferimento che si basa sulla manualistica consolidata (protocollo, testo, escatocollo e così via). Rispetto alla manualistica è stato aggiunto il "tenor-additum", una sorta di partizione multiuso, puramente funzionale alla risoluzione di problemi di annidamento delle sotto-partizioni tradizionali, utilizzato per codificare formule, clausole, pattuizioni

aggiuntive o altri elementi formali inseriti in posizioni diverse, che generano irregolarità nella struttura usuale.
– Funzioni personali giuridiche e documentarie. I marcatori identificano le funzioni 'reali' esercitate dalle persone nell'ambito dell'azione giuridica e della documentazione (autore, destinatario, giudice in una causa, testimone, notaio e via dicendo, con l'aggiunta di un marcatore denominato "res" per segnalare elenchi e descrizioni di beni e diritti oggetto dei negozi giuridici).
– Marcatura per l'indicizzazione automatica dei nomi di persona e di luogo e delle istituzioni ecclesiastiche, che vengono trattati previa regolarizzazione dei nomi medesimi (per luoghi e istituzioni è naturalmente prevista anche l'identificazione).
– Interventi editoriali. Questo livello di marcatura definisce le informazioni di carattere "tipografico" (sebbene, per esempio, si siano sostituite le parentesi quadre con l'inserimento di un colore diverso per il testo restituito); queste informazioni, che non vengono visualizzate, sono in realtà istruzioni per il *software*, che produrrà un *output* del documento in cui quelle istruzioni risulteranno eseguite.

Qui di seguito è riportato un testo sottoposto a codifica, ove sono cioè esplicitate tutte le cosiddette meta-informazioni che non vengono visualizzate e che riguardano invece i quattro livelli che abbiamo illustrato poc'anzi. In grassetto è segnalato il testo che viene effettivamente visualizzato (a meno naturalmente della formattazione, che viene definita in base alle istruzioni non visualizzabili).

Il *file* si apre con un blocco di istruzioni necessarie per indirizzarlo al foglio di stile e alla DTD.

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

Il secondo blocco, suddiviso in informazioni di carattere "editoriale" (<INFOED>) e per il *database* (<INFODB>) contiene informazioni relative a titolo, data e regesto del documento, tradizione e note introduttive, oltre ad alcuni dati che non vengono visualizzati, come il nome del *file*, gli estremi cronologici espressi in forma standardizzata e altre informazioni che servono a costruire un *database* di riferimento (che gestisce alcuni dati seriali quali gli estremi cronologici, il titolo, la segnatura, la provenienza etc.).

<EDITIO>
<INFOED>
<FILE>sam1181-06-24</FILE>

<AREA>**Pergamene milanesi**</AREA>
<FONDO>**S. Ambrogio**</FONDO>
<NUMERO>**7**</NUMERO>
<TIT-DOC>**Carta venditionis**</TIT-DOC>
<DATA>**1181 giugno 24, Comabbio.**</DATA>
<REG>**Giovanni, figlio del fu Giovanni** <TXT>**Ferrarius de Comabbio**</TXT>**, di legge longobarda, vende e cede a Giovanni detto** <TXT>**de Besozo**</TXT>**, monaco di S. Ambrogio, a nome della chiesa di S. Sepolcro** <INT>**di Ternate**</INT>**, un terreno con alberi che possiede nel territorio di Comabbio, ove dicesi** <TXT>**in Costa de Polegia**</TXT>**, e i diritti annessi, al prezzo di due soldi di moneta nuova e pone come fideiussore Alberto** <TXT>**de Castro Novo de Comabio**</TXT>.</REG>
<APPARATO>
<TRADITIO>**Originale, ASMi, AD, pergg., cart. 313, n. 199 [A]. Regesto del 1738 in Giorgi, Registro, c. 501; del 1739 in Giorgi, Rubrica, c. 27r.**</TRADITIO>
<P/>
<VERSO>**Nel verso, di mano trecentesca,** <TXT>**Vendit**<ABBR>**io**</ABBR></TXT>**; annotazione seicentesca di oggetto, data e segnatura** <TXT>**n. 106**</TXT>**; riferimenti all'Exemplaria Diplomatum del Giorgi; data di mano del Bonomi** <TXT>**MCLXXXI**</TXT>**.**</VERSO>
<P/><EDIZIONE></EDIZIONE>
<BIBLIO></BIBLIO>
<P/><OSSERVAZIONI>**Cattivo stato di conservazione, rosicature con perdita di testo nella parte sinistra, macchie diffuse. Tracce di rigatura.**</OSSERVAZIONI>
</APPARATO>
</INFOED>

<INFODB>
<FILE>sam1181-06-24</FILE>
<VALID>
<FROM>1181-06-24</FROM>
<TO>1181-06-24</TO>
</VALID>
<LOC>Comabbio</LOC>
<TIT>Carta venditionis</TIT>
<TRAD>Originale</TRAD>
<SEGN>
<ARCH>ASMi</ARCH>
<FONDO>Archivio Diplomatico – Pergamene per fondi</FONDO>
<PEZZO>313</PEZZO>

<ALTRO>n. 199</ALTRO>
</SEGN>
<ARCH-PROV>S. Ambrogio, monastero</ARCH-PROV>
<TERRITORIO>Milano</TERRITORIO>
<CONFECTIO>
<NOT>Alkerius Guaitamacus<QUAL>iudex</QUAL></NOT>
</CONFECTIO>
</INFODB>

Il terzo blocco è costituito dal testo del documento: qui, opportunamente annidati gli uni dentro gli altri, intervengono tutti i marcatori relativi ai quattro livelli di marcatura che costituiscono il cuore del trattamento elettronico del testo (strutture formali, partizioni e sotto-partizioni del documento, funzioni personali giuridiche e documentarie, marcatura per l'indicizzazione automatica dei nomi di persona e di luogo e delle istituzioni ecclesiastiche, interventi editoriali).

<TENOR><PROTOCOLLO><REST>**(SN)**</REST><DTCRON stl="natività"><REST> **Ann**</REST>**o dominice incar**<ABBR>**nacionis**</ABBR> **mill**<ABBR>**esimo**</ABBR> **octuagesimo primo, octavo kal**<ABBR>**endas**</ABBR> **iullii, indic**<ABBR>**ione**</ABBR> **.XIIII.**</DTCRON></PROTOCOLLO><TESTO><DISPOSITIO> **Car**<ABBR>**tam**</ABBR> **vendic**<ABBR>**ionis**</ABBR> **fecit et investivit et in** <LB/><REST> **s**</REST>**uu**<REST>**m**</REST> **locum posuit** <AUCT><PERSONA nm="Iohannes" pat="Iohannes Ferrarius de Commabio qd" lex="langobarda">**Iohannes, filius** <PERSONA nm="Iohannes Ferrarius de Commabio qd" fil="Iohannes">**quondam item Iohannis Ferrarii de** <TOP nm="Commabium" id="Comabbio, Va">**Commabio**</TOP></PERSONA>**, qui professus est vivere lege Longobarda**</PERSONA></AUCT>**, in** <RECIP><PERSONA nm="Iohannes qui dicitur de Besozo monachus Sancti Ambrosii">**do**<ABBR>**m**</ABBR>**no** <LB/><REST> **Iohanne**</REST> **qui dicitur de** <TOP nm="Besozum" id="Besozzo, Va">**Besozo**</TOP>**, monacho Sancti Ambr**<ABBR>**osii**</ABBR></PERSONA>**, ad partem** <ECCL id="S. Sepolcro, chiesa" top="San Sepolcro, Ternate, Va">**ecclesie Sancti Sepulcri**</ECCL></RECIP>**, pro precio solidorum duorum nove monete quos ab eo con**<LB/><REST>**fessus est se ac**</REST>**cepisse, de** <RES>**pecia una terre cum arboribus desuper habente quam habet in teritorio** <TOP nm="Commabium" id="Comabbio, Va">**Co**<ABBR>**m**</ABBR>**mabii**</TOP>**, ubi dicitur in Costa de Polegia - co**<LB/><REST>**heret**</REST> **a mane et a meridie** <PERSONA nm="illi Blancones">**Blanconum**</PERSONA>**, a ser**<ABBR>**o** </ABBR><ECCL id="S. Sepolcro, chiesa" top="San Sepolcro, Ternate, Va">**Sancti Sepulcri**</

ECCL></RES> **-,** <FORMULAE>**et de omni iure et accione quam habet in ea ut ipse do**<ABBR>**m**</ABBR>**nus Iohannes, ad partem ipsius ec**<LB/><REST>**clesi**</REST>**e, habeat et teneat et per omnia in suum locum sit et omne suum ius et accionem optineat.** </FORMULAE><CLAUSULAE>**Et promisit et guadiam dedit pro suo** <LB/><REST> **fa**</REST>**cto ab omni persona guarentandi iure et nominatim ab uxore, suis expensis; et posuit ei fideiussorem**<NOTA> **(a)** </NOTA><FID><PERSONA nm="Albertus de Castro Novo de Comabio">**Albertum de Castro Novo** <LB/><REST> **de**</REST><TOP nm="Commabium" id="Comabbio, Va"><REST> **C**</REST>**omabio**</TOP></PERSONA></FID>**, in pena dupli.**</CLAUSULAE></DISPOSITIO></TESTO>
<ESCATOCOLLO><DTTOP> **Actum Co**<ABBR>**m**</ABBR>**mabio.**</DTTOP>
<P/><IT><REST>**I**</REST>**nter**<REST>**fu**</REST>**e**<ABBR>**re**</ABBR> **ibi** <PERSONA nm="Enebertus de Bisucio">**Eneb**<ABBR>**e**</ABBR>**rtus de** <TOP nm="Bisucium" id="Besozzo, Va">**Bisucio**</TOP></PERSONA>**,** <PERSONA nm="Pinamons">**Pinamons**</PERSONA>**,** <PERSONA nm="Baregacius">**Baregacius**</PERSONA> **et** <PERSONA nm="Ugo Blanco de Commabio">**Ugo Blanco** <NOTA>**(b)**</NOTA> **de** <TOP nm="Commabium" id="Comabbio, Va">**Commabio**</TOP></PERSONA>**, testes.**</IT>
<P/><COMPLETIO><REST>**(SN)**</REST> **Ego** <SCRIPT><PERSONA nm="Alkerius Guaitamacus iudex">**Alkerius Guaitamacus iudex**</PERSONA></SCRIPT> **tradidi et scripsi.**
</COMPLETIO></ESCATOCOLLO></TENOR>

<NOTE>
<P/>**(a)** <TXT>**fideissorem**</TXT> **in A.**
<P/>**(b) Lettura incerta: segno abbreviativo generico sulla** <TXT>**-o**</TXT> **che potrebbe indurre a leggere** <TXT>**Blanco**<ABBR>**nus**</ABBR></TXT>
</NOTE>

Infine, il *file* si chiude con alcuni dati conclusivi relativi al nome di editore e codificatore, che vengono visualizzati, e alla dichiarazione di completamento del lavoro e la data di ultima revisione, non visualizzati.

<INFOCUR>
<EDIT>**Ada Grossi**</EDIT>
<ENCODING>**Ada Grossi**</ENCODING>
<STATUS>completo</STATUS>
<LR>2004-12-31</LR>

</INFOCUR>
</EDITIO>

Ricapitolando, il procedimento è il seguente:
– edizione critica tradizionale (a parte gli indici – che si possono comunque aggiungere per un *e-book* o per un'edizione a stampa, anche se essi risultano per molti versi superflui vista la possibilità di ricerche pressoché illimitate sui testi –);
– marcatura XML del testo: qui, come nella compilazione di un indice tradizionale, devono essere risolti i problemi di identificazione delle persone (per quanto possibile e con molta prudenza) e, soprattutto, dei luoghi; quest'ultimo aspetto è particolarmente rilevante, sia in sé, sia nell'ottica di una condivisione delle liste dei toponimi in un orizzonte più ampio (citiamo ancora "Lombardia Storica"[21]);
– caricamento dei *files* così prodotti sul *server*, ove vengono elaborati da un foglio di stile che vale per tutti i documenti caricati sullo stesso *server* e riconosce gli elementi formali della marcatura per restituire infine i *files* come pagine *web* visualizzate secondo convenzioni stabilite (anche il foglio di stile, come la DTD, può essere modificato o aggiornato in qualsiasi momento, dando luogo automaticamente alla nuova visualizzazione senza intervenire sui singoli *files* di edizione).

5. La "portabilità" dei dati

A proposito della produzione dei *files* che contengono i testi marcati, deve essere sottolineato che il tipo di trattamento elettronico dei testi adottato per il CDLM si distingue per quella che viene chiamata "portabilità dei dati". I *files* di codifica, in sé, sono dei semplici *files* di testo che hanno la caratteristica di essere "conservativi" (di essi, cioè, si garantisce la conservazione nel tempo). Infatti, la loro natura di *files* costruiti in solo testo comporta automaticamente che le informazioni che contengono non andranno perdute alla prossima tappa dell'evoluzione dei sistemi operativi, del *software* o del formato dei *files*. È possibile aprire e modificare gli stessi *files* con le macchine più obsolete e sarà sempre possibile aprirli e modificarli in futuro, finché ci saranno macchine in grado di leggere un *file* di testo. Potranno cambiare il foglio di stile, l'architettura del sito, perfino il sistema operativo: ma non cambieranno i *files* prodotti oggi, che saranno sempre utilizzabili.

6. Information retrieval

Infine, è giusto riservare qualche riferimento conclusivo al motore di ricerca TReSy (Text Retrieval System), che, come abbiamo già accennato, è un sistema

21 http://plain.unipv.it.

avanzato di ricerca automatica sui testi che offre enormi potenzialità.

Un testo scritto naturalmente non può rendere giustizia alla complessità delle ricerche, semplici e avanzate, che permette un simile strumento, pensato per la rete e non per la carta di queste pagine.

Nella speranza che chi legge abbia la curiosità di sperimentare qualche sessione di ricerca *on line* con il vero TReSy, diremo che, in generale, esso è in grado di scandagliare i testi in modo più o meno strutturato a seconda delle esigenze di chi lo utilizza: chi interroga TReSy stabilirà di volta in volta come filtrare gli ambiti d'interesse.

La ricerca più semplice, come sempre, è quella di una sequenza di caratteri lanciata in tutto il CDLM: ma è altresì possibile limitare il numero dei documenti che il motore di ricerca dovrà compulsare imponendo restrizioni di tipo cronologico, territoriale, tipologico, e/o circoscrivendo la ricerca ad alcune parti del documento in riferimento alla sua struttura formale e alle relative partizioni.

A una ricerca semplice può accedere chiunque, per cercare le attestazioni di un nome, di un luogo, di un lemma. Per una ricerca avanzata, invece, è necessaria la conoscenza della materia e un obiettivo più preciso: tanto per lanciare qualche suggestione, è possibile recuperare il testo corrispondente alle partizioni e alle sotto-partizioni formali dei documenti per studiarle sistematicamente, che siano le invocazioni verbali, le arenghe dei documenti pontifici o le sottoscrizioni dei notai; fare ricerche su nomi e funzioni personali per ricostruire carriere di giudici, di notai, di consoli o di qualunque altro personaggio; recuperare le clausole dei contratti per comprenderne l'evoluzione nel tempo nelle diverse aree; compiere studi comparati sugli usi cronologici; mappare i termini di scadenza dei fitti annuali; analizzare la lingua dei documenti; isolare citazioni giuridiche o scritturali.

L'elenco degli spunti di ricerca può e deve ampliarsi attraverso un utilizzo sempre più consapevole e sempre più diffuso tra gli studiosi delle discipline medievistiche.

Ada Grossi, Von der mittelalterlichen Schriftlichkeit zur digitalen Edition. Das Beispiel des Codice Diplomatico digitale della Lombardia Medievale (CDLM)

Der CDLM entstand 2000 an der Universität Pavia unter der Leitung von Michele Ansani mit Unterstützung der anderen Universitäten der Lombardei und unter der Schirmherrschaft der Region Lombardei. Der CDLM sieht die Totalerschließung, Edition und Digitalisierung der urkundlichen Überlieferung in kirchlichen, weltlichen sowie privaten Archiven der Lombardei von den Anfängen bis zum Ende des 12. Jahrhunderts vor. Das Projekt sah sich

zunächst einer Vielzahl von derzeit praktizierten Editionsverfahren gegenüber und hatte daher eigene einheitliche Grundsätze der Korpuserschließung zu entwickeln.

Neben der Bereitstellung größerer Textmengen zielt es auf deren vertiefte Erschließung mit verschiedenen Recherchemöglichkeiten ab. Die Indizierung erlaubt etwa die gezielte Suche nach Notaren und Zeitabschnitten, künftig auch von Faksimiles. Die automatisierten Suchfunktionen beruhen auf einem eigens von der Scuola Normale Superore di Pisa entwickelten Programm: Signum – Centro di ricerche informatiche per le discipline umanistiche (Zentrum für digitale Forschung im Bereich der Geisteswissenschaften).

Der Zugriff auf die Daten des CDLM und deren Recherche mit TreSy (*Text Retrieval System*) wird beispielhaft vorgeführt. Beim CDLM handelt es sich nicht um eine digitalisierte, sondern um eine digitale Edition. Es findet also keine einfache Übertragung herkömmlicher Editionen ins Netz statt, die Datenaufbereitung erfolgt vielmehr nach dynamischen Prinzipien. Der Textkorpus ist daher keine statische Größe, sondern ein mittels Markierungen und Hypertextfunktionen vielfältig vernetztes Datenmaterial, was dessen Benutzung enorm erleichtert.

Diese digitale Erschließung auf höherem Niveau eignet sich für mittelalterliche Quellentexte mit ihren oftmals stereotypierten Strukturen in hervorragender Weise. Urkunden und Akten des Mittelalters weisen hierarchische Muster auf, die der elektronischen Aufbereitung gut entgegenarbeiten. Die Ansätze der klassischen Diplomatik und Urkundenkritik können daher bestens in digitale Erschließungsvorhaben einfließen. Die für die vorgenommenen Markierungen verwendete Metasprache ist eine standardisierte XML-Anwendung (*eXtensible Mark-up Language*). Die einzelnen Hypertextfunktionen entsprechen dabei wissenschaftlichen Kriterien und richten sich an der Struktur der Urkundentexte selbst aus. Die gebotenen, nicht unmittelbar sichtbaren Metainformationen betreffen die Struktur der Texte, seine Teile, die rechtlichen Funktionen von Urkunden, die in ihnen enthaltenen Personen- und Ortsnamen sowie kirchlichen Organisationen, nicht zuletzt die von den Redakteuren vorgenommenen Eingriffe und Emendationen. Insbesondere werden alle Toponyme sowie Namen von Personen oder Organisationen automatisch in Registerlisten überführt. Ein weiterer Vorteil der Datenbankstruktur ist die Handlichkeit der in ihr verarbeiteten Textsorten. Die einzelnen Dateien sind einfache Textfiles, was deren Langzeitarchivierung am ehesten gewährleistet. Besonderes Augenmerk wird schließlich auf die verschiedenen Ebenen der Markierung von Urkundentexten gelegt, ob dies nun deren formale Struktur, deren Abschnitte und Unterabschnitte, die Rechtsfunktionen der Aktanten, die automatische Indizierung von Personen, Orten und geistlichen Anstalten oder einzelne editorische Maßnahmen betrifft.