Sulla strada dell'eXtensible Markup Language

Un confronto tra i modelli MASTER e MNSS

La catalogazione dei manoscritti, spesso considerata un'attività di nicchia, si sta affiancando concretamente alle tecnologie informatiche su una delle vie più significative della nostra epoca, quella dell'eXtensible Markup Language.[1] Questo linguaggio, che costituisce un ponte tra la produzione delle informazioni e il grande archivio pubblico costituito dal web, ha numerosi pregi. Innanzi tutto, per volontà esplicita del gruppo di lavoro del World Wide Web Consortium, che è responsabile della creazione e del mantenimento dello standard, un obiettivo di XML è risultare "umanamente leggibile e ragionevolmente chiaro", e di fatto esso è facilmente comprensibile anche a coloro che non possiedono una particolare preparazione tecnica. In secondo luogo XML permette l'elaborazione di singoli standard applicabili all'interno di specifici contesti di ricerca e di lavoro.[2]

Nel campo della catalogazione dei manoscritti letterari e storici esistono attualmente proposte diverse che sfruttano questo formato.

Il progetto Manuscripts Access through Standard for Electronic Records, sovvenzionato dalla Comunità europea, ha recentemente portato a termine il proprio mandato lasciando in eredità una Document Type Definition (DTD), disponibile in web e accompagnata da un approfondito *Reference manual*.[3]

Il Medieval Manuscripts Description Work Group della Text Encoding Initiative, con cui i partecipanti a MASTER hanno lungamente collaborato, sta ora elaborando una proposta parzialmente diversa, più flessibile e più ampia. Essa è visibile pubblicamente nelle minute degli ultimi incontri del gruppo.[4]

Comune al modello di MASTER e quello sostenuto dal MMSS Group di TEI è l'istituzione dell'elemento <msDescription>, destinato a contenere la descrizione di un singolo documento manoscritto. Per entrambi i progetti tale elemento dovrebbe costituire un'estensione dell'attuale TEI DTD, finalizzata all'edizione di testi letterari di contenuto umanistico.[5] In quanto tale <msDescription> può comparire o all'interno del <teiHeader> oppure annidato all'interno di <text> e di <body>, i marcatori che introducono il contenuto vero e proprio di un documento elettronico TEI compatibile. All'interno del <teiHeader> il tag <msDescription> contiene la descrizione della fonte documentaria da cui proviene il testo pubblicato nel corpo del documento; all'interno di <text> e i <body> il tag <msDescription> costituisce il contenuto specifico del documento TEI.

Le differenze più rilevanti tra il modello di MASTER e il modello del MMSS Group, che si rilevano soprattutto ai piani alti delle loro DTD, riguardano in primo luogo i *children* diretti di <msDescription>. MASTER descrive questo elemento con la dichiarazione

<!ELEMENT msDescription (msIdentifier, msHeading?, msContents?, physDesc?, history?, additional?, msPart*)>[6]

Il MMSS Group descrive l'elemento <msDescription> con la dichiarazione

<!ELEMENT msDescription (msIdentifier, msHeading?, (msContents | physDesc | history | additional | msPart)*)>[7]

Sia per MASTER sia per il MMSS Group la descrizione di un manoscritto deve obbligatoriamente contenere all'inizio un tag <msIdentifier>, destinato a marcare le informazioni che permettono di identificare un singolo manoscritto, tradizionalmente costituite dalla città, dal nome della biblioteca e dalla segnatura. Quindi, entrambi i modelli prevedono l'inserimento del tag <msHeading>, non obbligatorio e non ripetibile, che comprende la descrizione sommaria del manoscritto.

Le differenze fondamentali tra le due dichiarazioni riguardano invece l'ordine e la ripetibilità degli altri tag che possono apparire in <msDescription>, ossia i marcatori destinati a contenere la descrizione interna, <msContents>, la descrizione esterna, <physDesc>, la storia del documento, <history>, le notizie gestionali e la bibliografia, <additional>, la descrizione di eventuali porzioni del manoscritto, ad esempio quando esso sia composito, <msPart>.

Questi cinque elementi, secondo lo standard proposto da MASTER, non sono obbligatori, non sono ripetibili più di una volta (eccetto ovviamente <msPart>), e devono essere usati solo nell'ordine in cui appaiono nella dichiarazione.

Anche lo standard studiato dal MMSS Group prevede che nessuno di questi tag sia obbligatorio, ma permette di ripeterli qualsiasi numero di volte e di usarli in ordine libero, purché vengano di seguito a <msIdentiifer> e a <msHeading>.

Altre differenze significative tra le due DTD consistono nell'uso dei tag all'interno di alcuni dei *children* di <msDescription>.

Fig. 1 - Ms. Casanatense 716, descrizione interna e storia, modello di MASTER

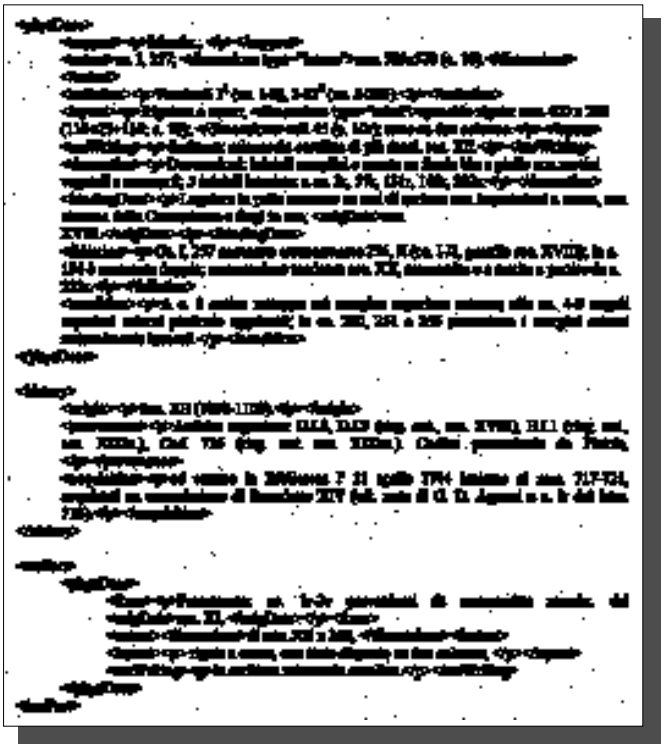
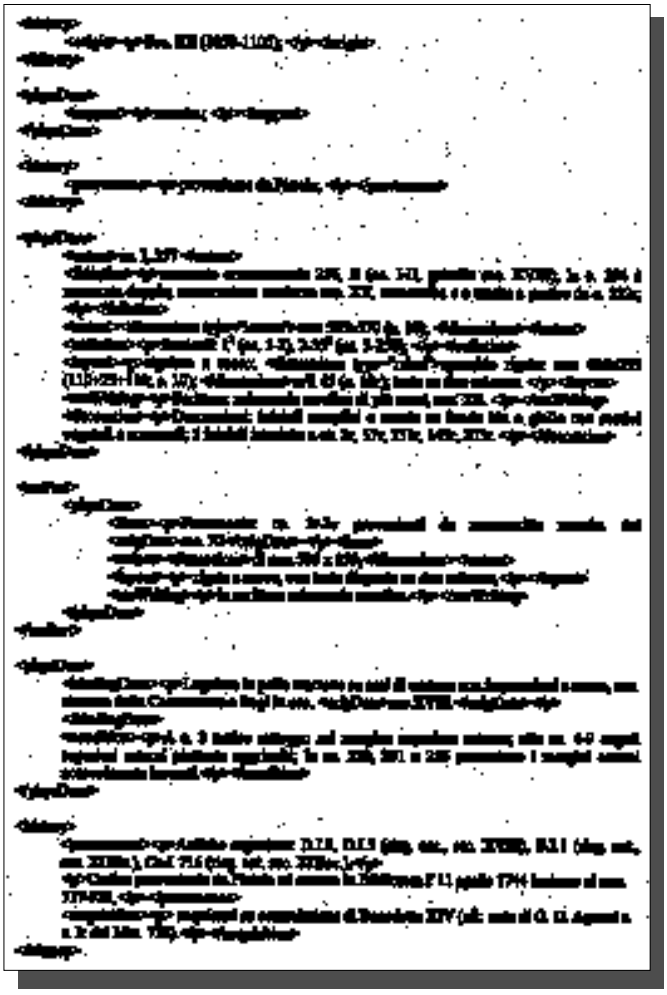


Fig. 2 - Ms. Casanatense 716, descrizione interna e storia, modello di MMSS Group



MASTER dichiara <msContents>, <physDesc> e <history> come segue:

<!ELEMENT msContents (p+ | (overview?, msItem+))>

<!ELEMENT physDesc (p+ | (form?, support?, extent?, collation?, layout?, msWriting?, musicNotation?, decoration?, bindingDesc?, foliation*, additions?, condition?))>

<!ELEMENT history (p+ | (origin?, provenance*, acquisition?))>[8]

Il MMSS Group di TEI propone invece:

<!ELEMENT msContents (p | msItem)+>

<!ELEMENT physDesc (p | form | support | extent | collation | layout | msWriting | writingSystem | musicNotation | decoration | bindingDesc | foliation | additions | condition | signatures | catchwords) +>

<!ELEMENT history (p | origin | provenance | acquisition) +>[9]

Lo schema del MMSS Group favorisce la codifica di schede catalografiche anche molto diverse tra loro, sia che si tratti di schede di nuova produzione, sia che si tratti di descrizioni recuperate da progetti precedenti, i cui criteri di catalogazione non risultano più discutibili. MASTER, invece, propone al catalogatore di adeguare i propri criteri a uno schema più rigido, oppure, nel caso si debbano recuperare informazioni precedenti, il progetto europeo richiede di modificare, all'interno dei documenti XML, la disposizione delle informazioni presenti nella fonte.

In aggiunta si noti che MASTER permette l'uso del tag <p> in alternativa all'uso degli altri tag, ma non ne permette l'inserimento parallelo allo stesso livello, sia in <msContents>, sia in <physDesc> e in <history>. TEI MMSS, invece, in queste stesse aree descrittive permette allo stesso livello l'uso di tutti i tag, compreso l'elemento <p>, ripetuti qualsiasi numero di volte, in qualsiasi ordine, e lascia così la libertà di inserire porzioni di prosa in qualsiasi punto della descrizione.[10]

Un esempio potrà aiutare a comprendere meglio le differenze fin qui esposte e a considerarne il valore in rapporto al contesto italiano. Partiamo dalla descrizione interna del manoscritto Roma, Biblioteca Casanatense 716, inserita tra gli esempi nelle *Norme per la descrizione uniforme dei manoscritti in alfabeto latino,* pubblicate a cura dell'Istituto centrale per il catalogo unico e della Commissione indici e cataloghi.

Il testo, copiato fedelmente dalla fonte, è il seguente:

> Sec. XII (1050-1100); membr.; proveniente da Pistoia; cc. I, 257 numerate erroneamente 256, II (cc. I-II, guardie sec. XVIII); la c. 194 è numerata doppia; numerazione moderna sec. XX, meccanica e a matita a partire da c. 222r; mm 585x370 (c. 10); fascicoli: 1^2 (cc. 1-2), $2\text{-}33^8$ (cc. 3-256); rigatura a secco; specchio rigato: mm 400 x 235 (110+25+110; c. 10); rr/ll 45 (c. 10r); testo su due colonne.
>
> Scrittura: minuscola carolina di più mani, sec. XII.
>
> Decorazioni: iniziali semplici e ornate su fondo blu e giallo con motivi vegetali e zoomorfi; 5 iniziali istoriate: a cc. 3r, 57r, 131r, 145r, 203r.
>
> Frammento: cc. 1r-2v provenienti da manoscritto membr. del sec. XI, di mm 506 x 359, rigate a secco, con testo disposto su due colonne, in scrittura minuscola carolina.
>
> Legatura in pelle marrone su assi di cartone con impressioni a secco, con stemma della Casanatense e fregi in oro, sec. XVIII.
>
> A c. 3 antico rattoppo sul margine superiore esterno; alle cc. 4-9 angoli superiori esterni piuttosto aggrinziti; le cc. 250, 251 e 256 presentano i margini esterni notevolmente lacerati.
>
> Antiche segnature: D.I.8, D.I.3 (risg. ant., sec. XVIII), B.I.1 (risg. ant., sec. XIXin.), *Cod.* 716 (risg. ant., sec. XIXex.).
>
> Codice proveniente da Pistoia ed entrato in Biblioteca l' 11 aprile 1744 insieme ai mss. 717-721, acquistati su commissione di Benedetto XIV (cfr. nota di G. D. Agnani a c. Ir del Mss. 718).[11]

Qui le notizie fisiche e quelle relative alla storia del codice sono unite in un'unica area descrittiva, secondo una prassi spesso seguita tanto a livello bibliotecario quanto a livello accademico e di ricerca.

Nella figura 1 si può vedere la descrizione interna del manoscritto Casanatense codificata in XML secondo la DTD proposta da MASTER, mentre nella figura 2 tale descrizione ricompare nella forma valida per il MMSS Group. Poiché lo scopo dei due esempi è mostrare le differenze delle due DTD ai loro livelli più alti, all'interno dei tag principali è stata usata una marcatura valida, ma di tipo generico.

Il modello proposto da MASTER (figura 1) ha richiesto di mutare l'ordine originario delle informazioni. La datazione ["Sec. XII (1050-1100)"] è stata spostata dall'*incipit* della scheda al tag <msHistory>, che secondo la DTD deve essere posto dopo <physDesc>; la numerazione delle carte ["cc. I, 257 numerate ... a partire da c. 222r;"],

che costituisce il contenuto del tag <foliation>, è slittata dopo la descrizione della legatura; la descrizione del frammento interno ["Frammento... minuscola carolina."] ha trovato posto alla fine della scheda, poiché l'elemento <msPart> deve seguire tutti gli altri *children* di <msDescription>.

Il modello proposto da MMSS Group ha permesso di lasciare le informazioni nell'ordine originario, aprendo e chiudendo più volte i tag <history> e <physDesc>; anche le notizie sul frammento sono state inserite nel punto in cui si trovavano nella scheda originale, poiché il tag <msPart>, secondo questa DTD, può essere ripetuto in qualsiasi punto della descrizione.

Le due DTD, che attualmente propongono due soluzioni diverse, non sono comunque completamente alternative tra loro, poiché la prima, quella del gruppo europeo, costituisce un sottoinsieme della seconda ed è con essa compatibile. Ciò significa che un documento valido per MASTER lo è necessariamente anche per il MMSS Group, mentre non tutti i documenti validi per il MMSS Group sono validi per MASTER.

La valutazione degli schemi proposti è ora lasciata al giudizio del pubblico. Poiché nessuna delle DTD studiate ha ricevuto riconoscimenti formali, neanche da parte del Consorzio TEI, avviarsi ora sulla strada di XML ha un duplice valore: seguire da testimoni diretti la nascita di uno standard e avere la possibilità di intervenire su di esso. Strada facendo, in questo modo, sarà possibile acquisire la consapevolezza, altrimenti difficilmente recuperabile, delle ragioni che determineranno la forma definitiva di queste DTD.

Note

[1] Un avvio alla conoscenza di XML e di alcuni elementi del linguaggio, che darò qui per scontato, si legge in V. Pasqui, *XML: un'introduzione,* "Biblioteche oggi", 19 (2001), 6, p. 18-24.

[2] Gli scopi di XML sono esplicitamente trattati in T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, *Extensible Markup Language (XML) 1.0 (second edition),* W3C Reccomendation 6 October 2000, paragrafo 1.1, <http://www.w3.org/TR/REC-xml>.

[3] Manuscript Access through Standard for Electronic Records (d'ora in avanti MASTER), <http://www.cta.dmu.ac.uk/projects/master>.

[4] Text Encoding Initiative Medieval Manuscripts description (d'ora in avanti TEI MMSS) Work Group, <http://www.stg.brown.edu/~tei-mmss/index.html>. Si veda anche il progetto Codices Electronici Ecclesiae Coloniensis (CEEC) dell'Università di Colonia, che pur rifacendosi esplicitamente ad alcune delle direttive di MASTER e di TEI, ha elaborato una DTD particolare e più complessa, pubblicata all'indirizzo <http://ceec.uni-koeln.de>.

[5] Cfr. <http://www.tei-c.org/Guidelines2/index.html>.

[6] La DTD completa di MASTER, qui citata nella versione del 19 febbraio 2001, è disponibile all'indirizzo <http://www.hcu.ox.ac.uk/TEI/Master/Reference/DTD/masterx.dtd>.

[7] La dichiarazione è pubblicamente disponibile nella minuta dell'incontro del MMSS Group risalente al 30 settembre – 1 ottobre 2000, all'indirizzo <http:// www.stg.brown.edu/~tei-mmss/collegepark.html>.

[8] Cfr. <http://www.hcu.ox.ac.uk/TEI/Master/Reference/DTD/masterx.dtd>.

[9] Cfr. <http:// www.stg.brown.edu/~tei-mmss/collegepark.html>.

[10] Altre differenze, riguardanti singoli tag e singoli attributi, potranno essere dedotte dal confronto diretto delle pagine web citate alle note 5 e 6.

[11] *Norme per la descrizione uniforme dei manoscritti in alfabeto latino,* a cura del Ministero per i beni e le attività culturali, Ufficio centrale per i beni librari e le istituzioni culturali e l'editoria, Istituto centrale per il catalogo unico e per le informazioni bibliografiche, Commissione indici e cataloghi, Roma, ICCU, 2000, p. 37-38.