

## La struttura della fonte e la sua resa digitale: la DTD

Fruire in modo non superficiale delle potenzialità offerte dall'automazione di alcuni procedimenti implica, quale presupposto metodologico, la conoscenza puntuale quanto approfondita del contenuto e della struttura dei dati codificati e assieme, dei criteri che hanno presieduto la loro costruzione. Come è più volte emerso nei paragrafi precedenti, l'operazione fondamentale da compiere per poter utilizzare la codifica informatica nel campo storico

è quella di individuare e descrivere in modo non ambiguo la procedura critica a partire dalla definizione degli oggetti, delle unità minime, delle entità e delle relazioni che animano l'obiettivo del nostro studio. Oggetti, unità minime, entità e relazioni che fanno parte integrante della strumentazione – e quindi del metodo – del critico<sup>1</sup>.

In questa direzione, l'*eXtensible Markup Language* è in grado di fornire uno strumento utile e adeguato: la *Document Type Definition (DTD)*<sup>2</sup>. Una definizione, in via preliminare: una *Document Type Definition* è un insieme di regole per costruire documenti *XML* sintatticamente validi (*well formed*)<sup>3</sup>, essenzialmente una grammatica per la descrizione dei *tag* e del loro innestamento. Di più: una *DTD* definisce l'insieme di tutte le possibili marcature valide per un documento e in essa vengono stabiliti, a priori, tutti gli elementi che si ritengono significativi della struttura e il contenuto dei documenti, ma anche i loro attributi e i rapporti tra essi intercorrenti. Una *DTD* è dunque un file<sup>4</sup> che definisce nomi, attributi e restrizioni di occorrenza simultanea per tutti gli elementi e le entità identificate e usate da un'intera classe di documenti, stabilendone la forma e le reciproche relazioni. In concreto essa è costituita da un insieme di dichiarazioni di *markup* definite da un nome (identificatore generico) e un modello di contenuti (*content model*), divise in:

---

<sup>1</sup> G. GIGLIOZZI, *La galassia Von Neumann: il testo tra piombo e byte*, in *I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici* cit., pp. 209-231:228; disponibile on line all'indirizzo: <http://crllet.scu.uniroma1.it/pubblicazioni/saggi/Lincei98.htm>.

<sup>2</sup> Lo stesso linguaggio *XML* ha una propria *DTD*, attualmente descritto nella specifica REC-xml-19980210, in cui vengono elencate le regole della specifica stessa del linguaggio: le specifiche sono scaricabili all'indirizzo: <http://www.w3.org/TR/1998/REC-xml-19980210>.

<sup>3</sup> Un documento *XML* è ben formato – e quindi analizzabile da un programma (*parser*) - se e solo se: tutti i suoi *tag* sono chiusi, nell'ordine corretto; esiste un solo elemento radice; i valori degli attributi sono scritti tra virgolette; gli elementi non hanno due attributi con lo stesso nome; i nomi degli elementi e degli attributi non contengono i caratteri < >.

<sup>4</sup> Di solito i file contenenti una *DTD* sono esterni al documento *XML* (*external set*) e vengono richiamati da questi mediante un percorso posto in apertura, subito sotto la dichiarazione *XML*. Tuttavia non sono rari i casi in cui la *DTD* si annidi all'interno dello stesso documento *XML* (*internal set*), senza per questo inficiarne la validità. Questo sistema comporta però uno svantaggio in termini di condivisione in quanto la *DTD* non può essere usata per altri documenti *XML* simili; viceversa, mantenendo una *DTD* esterna, questa risulterà condivisibile e riutilizzabile.

- dichiarazioni dei tipi di elemento,
- dichiarazioni degli attributi associati ad un dato elemento,
- dichiarazioni delle entità (*entity*)<sup>5</sup>.

L'elemento identifica un'unità testuale considerata come componente strutturale: in questo senso, contenendo un nucleo informativo, è in qualche modo simile ad un campo di *database* anche se, diversamente da questo, mostra una flessibilità ben maggiore, rendendo possibile allargare o specificare l'informazione in esso contenuta. Ad ogni elemento viene associato un *content model* che specifichi cosa possa essere contenuto al suo interno: altri elementi, attributi o testo libero (`#PCDATA`) attraverso indicatori di occorrenza deputati a stabilire quante volte quel determinato contenuto possa essere utilizzato (\* = zero o più occorrenze, ? = zero o 1 occorrenza, + = una o più occorrenze); l'assenza di indicatori indica una e una sola occorrenza. Un esempio, molto semplice, di occorrenze di elementi in una *Document Type Definition* può essere il seguente:

```
<!ELEMENT CAPITOLO (titolo, sottotitolo?, paragrafo+)>
<!ELEMENT PARAGRAFO (#PCDATA | note)*>
<!ELEMENT note (#PCDATA)>
```

In questo modello, all'interno dell'elemento capitolo sono previsti un titolo, qualora presente un unico sottotitolo, mentre si stabilisce che l'elemento paragrafo debba essere presente almeno una volta. L'elemento note invece, contenuto all'interno del paragrafo, può essere ripetuto infinite volte oppure essere totalmente assente.

Gli attributi associati ad ogni elemento vengono esplicitati attraverso una lista introdotta dalla dichiarazione `<!ATTLIST...>` nella quale il singolo attributo viene definito da un nome, il tipo di dato e un valore di *default*, stabilito tra `#REQUIRED` (quando il valore deve essere specificato) e `#IMPLIED` (quando non deve essere specificato nessun particolare valore):

---

<sup>5</sup> Non sembra questa, la sede adatta per discutere il ruolo delle entità, che meriterebbero una trattazione lunga e autonoma. Basti qui sapere che questo genere di dichiarazioni riguardano parametri che possono essere esterni, e pertanto utilizzabili ad esempio per inserire file all'interno di un documento *XML*, o interni. In questo le entità vengono utilizzate per la rappresentazione di singoli caratteri non previsti nel *set XML* tramite il loro numero di codice (esprimibile con la forma: "&#num;," dove num è un numero di ASCII/Latin-1/UNICODE): ad esempio `<!ENTITY apos " &#39 ; ">` sostituisce con il codice l'apostrofo, normalmente utilizzato da *XML* e quindi in conflitto.

<!ATTLIST edizione

Iden ID #IMPLIED

Status (bozza | rivisto | pubblicato) #REQUIRED

lingua CDATA "italiano">

Nel caso illustrato si stabilisce ad esempio che nell'inserire l'attributo edizione il codificatore debba sempre segnalare se si tratta di una bozza, di un testo rivisto oppure di un testo già pubblicato.

Già da questi esempi basilari si intuisce quella che può essere la centralità di una *Document Type Definition* in una procedura di codifica informatica, vista la sua indubbia utilità nell'orientare e vincolare l'interpretazione della fonte analizzata, sia per quanto concerne il contenuto che si intende comunicare, sia con riferimento alla predisposizione di un quadro di riferimento utile all'utente che si approcci al testo codificato. Tra gli indubbi vantaggi insiti nell'utilizzo di una *DTD* nella composizione di un'edizione elettronica, vanno infatti annoverati la possibilità – per ogni documento edito – di contenere al suo interno una descrizione del proprio formato, ma anche quella – per gruppi di utenti diversi – di concordare un modello unitario di codifica che faciliti lo scambio della documentazione. Avendo inoltre predisposto preventivamente una *Document Type Definition*, qualunque *editor* deputato alla produzione di documenti *XML* è in grado di sfruttare le sue regole per la creazione di un'interfaccia di inserimento dei dati che sia conforme con il modello, evitando in questo modo eventuali errori di inserimento di marcatori da parte del codificatore. Resta per inteso che, dal momento che ogni definizione di tipologia documentaria, ogni modellizzazione, si delinea come interpretazione soggettiva, non esiste una *Document Type Definition* in grado di riflettere una verità assoluta del testo trattato. La *DTD*, in sostanza, è uno strumento che non fornisce informazioni sulla semantica né tantomeno sulla rappresentazione di un documento, ma il cui scopo è esercitare un controllo sulla marcatura: una guida, il cui uso assicura che tutti i documenti stabiliti siano informaticamente costruiti e strutturati in maniera omogenea e conforme.

Nelle pagine che seguono viene riportata dal *Document Type Definition* stabilita per i documenti del *Liber Privilegiorum* di Monreale. Anche in questo caso si è ritenuto di poter precisare e definire un modello autonomo, calibrato sulle specificità del materiale documentario utilizzato, scartando l'impiego della *DTD* proposta dalla *Text Encoding*

*Initiative*, sia nella versione completa sia in quella leggera che, come si è visto nel paragrafo precedente, sono orientate ai testi letterari, e quella messa a punto nell'ambito del progetto *MEP (Model Editions Partnership. Historical Editions in the Digital Age)* dell'University of South Carolina, mirato a formalizzare uno standard per la pubblicazione e la conversione di edizioni a stampa in formato elettronico di documentazione storica americana di età moderna<sup>6</sup>. Lo schema proposto rispecchia fedelmente gli interessi della ricerca in atto, che sono tutti orientati al contenuto del cartulario e dei documenti in esso trascritti, escludendo pertanto la codifica degli aspetti materiali del manoscritto, per i quali si è pensato di riservare all'interno del sito un'apposita sezione illustrativa. L'intento, non esclude tuttavia che in un secondo momento, se necessario, si possa procedere ad integrazioni della *DTD*, in grado di rendere la codifica testuale attuata conforme alle specifiche *TEI-CEI* o alle loro estensioni modulari.

---

<sup>6</sup> Cfr. <http://mep.cla.sc.edu/>.

<!--

=====  
DTD dei documenti del Liber Privilegiorum Sanctae Montis Regalis Ecclesiae nella tradizione del  
Codice Vaticano latino 3880

<http://www.vatlat3880.altervista.org>

Ultimo aggiornamento: 21/10/2008

=====  
-->

<!--Prologo obbligatorio all'inizio di ciascun documento: la seconda riga contiene il percorso al  
foglio di stile per la trasformazione in html-->

<?xml version="1.0" encoding="UTF-8" standalone="no"?>  
<?xml-stylesheet type="text/xsl" href="trasformazione.xsl"?>

<!ELEMENT EDITIO (NUMDOC, DOCPREC, DOCSUCC, INFOED, TENOR, PROTOCOLLO,  
TESTO, ESCATOCOLLO, NOTE)>

<!ELEMENT NUMDOC (#PCDATA)>

<!ELEMENT DOCPREC (#PCDATA)>

<!ELEMENT DOCSUCC (#PCDATA)>

<!ELEMENT INFOED (DATA, NUMCARTE, APPARATO)>

<!ELEMENT DATA (#PCDATA)>

<!ATTLIST DATA

luogo CDATA #IMPLIED

luogo (non-indicato) "non-indicato"

anno CDATA #IMPLIED

anno (non-indicato) "non-indicato"

mese CDATA #IMPLIED

mese (non-indicato) "non-indicato"

giorno CDATA #IMPLIED

giorno (non-indicato) "non-indicato"

ind CDATA #IMPLIED

ind (non-indicato) "non-indicato"

>

<!ELEMENT NUMCARTE (#PCDATA)>

<!ATTLIST NUMCARTE

cc CDATA #IMPLIED>

<!--esempio: NUMCARTE cc="1v A-3v A"-->

<!ELEMENT APPARATO (TRADITIO, ED, REG, BIBLIOGRAFIA, REGESTO)>

<!ELEMENT TRADITIO (ORIGIN, COP)>

<!ELEMENT ORIG (#PCDATA)>

<!ATTLIST ORIG

entecons CDATA #IMPLIED

fondo CDATA #IMPLIED

segn CDATA #IMPLIED

dim CDATA #IMPLIED

>

<!--esempio: ORIGIN entecons="BCRS" fondo="FM" segn="Balsamo 29" dim="mm. 850+36590"-->

<!ELEMENT COP (#PCDATA)>

<!ATTLIST COP

entecons CDATA #IMPLIED

segn CDATA #IMPLIED

>

<!--esempio: COP entecons="BSEM" segn="XX.E.8"-->

<!ELEMENT ED (#PCDATA)>

<!ATTLIST ED

editio1 CDATA #IMPLIED

editio2 CDATA #IMPLIED

editio3

>

<!--esempio: ED editio1="Pirro, Sicilia Sacra, I, p. 453" editio2="Del Giudice, Descrizione, p.2" editio3="Lo Monaco Aprile, Le Decime, pp.8-16"-->

<!ELEMENT REG (#PCDATA)>

<!ATTLIST REG

regesto1 CDATA #IMPLIED

regesto2 CDATA #IMPLIED

regesto3 CDATA #IMPLIED

>

<!--esempio: REG regesto1="Behring, Regesten, 2, p. 201" regesto2="Garufi, Catalogo, n.15, pp.10-11"-->

<!ELEMENT BIBLIOGRAFIA (#PCDATA)>

<!ATTLIST BIBLIOGRAFIA

cit1 CDATA #IMPLIED

cit2 CDATA #IMPLIED

cit3 CDATA #IMPLIED

>

<!--esempio: BIBLIOGRAFIA cit1="Barberi, Beneficia, I, p.34" cit2="Celauro, Il territorio, p.26" cit3="Concetti, Memorie, p.115ss."-->

<!ELEMENT REGESTO (#PCDATA)>

<!ELEMENT TENOR (#PCDATA | PERSONA | ECCL | TOP | BENIMM)\*>

<!ELEMENT PROTOCOLLO (INVOCATIO, INTITULATIO, INSCRIPTIO, DTCRON, DTTOP, APPRECATIO)>

<!ELEMENT INVOCATIO (#PCDATA)>

<!ELEMENT INTITULATIO (#PCDATA | PERSONA | ECCL | TOP)\*>

<!ELEMENT INSCRIPTIO (#PCDATA | PERSONA | ECCL | TOP)\*>

<!ELEMENT DTCRON (#PCDATA)>

<!ATTLIST DTCRON

anno CDATA #IMPLIED

anno (non-indicato) "non-indicato"

mese CDATA #IMPLIED

mese (non-indicato) "non-indicato"

giorno CDATA #IMPLIED

giorno (non-indicato) "non-indicato"

ind CDATA #IMPLIED

ind (non-indicato) "non-indicato"

>

<!ELEMENT DTTOP (#PCDATA | TOP)>

<!ELEMENT APPRECATIO (#PCDATA)>

<!ELEMENT TESTO (ARENKA, NARRATIO, DISPOSITIO, SANCTIO, CORROBORATIO)>

```

<!ELEMENT ARENGA (#PCDATA | PERSONA | ECCL | BENIMM | TOP)*>
<!ELEMENT NARRATIO (#PCDATA | PERSONA | ECCL | BENIMM | TOP)*>
<!ELEMENT DISPOSITIO (#PCDATA | PERSONA | ECCL | BENIMM | TOP)*>
<!ELEMENT SANCTIO (#PCDATA | PERSONA | ECCL | BENIMM | TOP)*>
<!ELEMENT CORROBORATIO (#PCDATA | PERSONA | SCRIPT| ECCL | BENIMM | TOP)*>
<!ELEMENT ESCATOCOLLO (#PCDATA, ROTA, DTTOP, RECOGNITIO, DTCRON,
SUBSCRIPTIO)>
<!ELEMENT ROTA (#PCDATA)
<!ATTLIST ROTA
    legendaest CDATA #IMPLIED
    legendaint CDATA #IMPLIED
    citbib CDATA #IMPLIED>
    <!--esempio: ROTA legendaest="Dextera Domini fecit virtutem. Dextera Domini exaltavit
me" legendaint="WILLELMUS DIVINA FAVENTE CLEMENCIA REX SICILIE, DUCATUS APULIE
ET PRINCIPATUS CAPUE." citbib="Ps 117(118), 16"-->
<!ELEMENT DTTOP (#PCDATA | TOP)>
<!ELEMENT RECOGNITIO (#PCDATA | PERSONA | TOP | ECCL)*>
<!ELEMENT DTCRON (#PCDATA)>
<!ELEMENT SUBSCRIPTIO (#PCDATA | SMS)*>
<!ELEMENT SMS (#PCDATA | PERSONA)*>
<!ELEMENT NOTE (#PCDATA | nota)*>

<!ELEMENT PERSONA (#PCDATA)>
<!ATTLIST PERSONA
    nm CDATA #REQUIRED
    nameattr CDATA #IMPLIED
    id CDATA #IMPLIED
    id (non-identificato) "non-identificato"
    tit CDATA #IMPLIED
    pat CDATA #IMPLIED
    mat CDATA #IMPLIED
    fil CDATA #IMPLIED
    vir CDATA #IMPLIED

```

ux CDATA #IMPLIED

germ CDATA #IMPLIED

fr CDATA #IMPLIED

sor CDATA #IMPLIED

par CDATA #IMPLIED

>

<!ELEMENT ECCL (#PCDATA)>

<!ATTLIST ECCL

nm CDATA #REQUIRED

id CDATA #IMPLIED

id (non-identificato) "non-identificato"

tipo (chiesa | monastero | arcivescovato | vescovato | abbazia)

ub CDATA #IMPLIED

ub (incerto) "incerto"

>

<!ELEMENT BENIMM (#PCDATA)>

<!ATTLIST BENIMM

nm CDATA #REQUIRED

id CDATA #IMPLIED

>

<!ELEMENT TOP (#PCDATA)>

<!ATTLIST TOP

nm CDATA #REQUIRED

id CDATA #IMPLIED

id (non-identificato) "non-identificato"

ub CDATA #IMPLIED

ub (incerto) "incerto"

tipo (toponimo | microtoponimo | el geografico)

subtipo CDATA #IMPLIED

>

<!-- Fanno parte dei toponimi i coronimi, ovvero designazioni ampie, quali regioni sia istituzionali che storiche, le regioni geografiche e i raggruppamenti; e le città -->

<!-- La microtoponomastica riguarda i nomi dei piccoli insediamenti abitativi e delle case isolate, i toponimi urbani o gli odonomi. Nella categoria "microtoponimo" rientrano: casale, castello, cimitero, coltivazione, cortile, croce, fortificazione, giardino, guado, mulino, porta, porto, recinto, scala, torre, via, vigna -->

<!-- Nella categoria "el geografico" rientrano: campo, canale, collina, conca, fiume, fonte, fossa, gorgo, grotta, isola, lago, monte, palude, pietra, prato, rocca, rupe, sorgente, valle -->

<!ELEMENT SCRIPT (#PCDATA)>