

I metodi per la conservazione: il ruolo dei metadati e XML

Maria Guercio (Università di Urbino)

Premessa

Conservare gli oggetti della memoria è un'attività impegnativa e costosa che gli individui e le organizzazioni tendono progressivamente e inesorabilmente a considerare in termini residuali. L'accorciamento dell'orizzonte temporale è un fenomeno generale e ha conseguenze rilevanti proprio in questo campo tanto da determinare atteggiamenti diffusi e preoccupanti di sufficienza e impazienza soprattutto per quella parte del patrimonio culturale che non è immediatamente e facilmente riconoscibile e fruibile.

Gli archivi rientrano sicuramente in questa categoria, anche se la normativa in materia, recentemente aggiornata, considera anche la documentazione corrente e di deposito come beni culturali, oggetto perciò di tutela rigorosa da parte dello Stato italiano, al punto da introdurre sanzioni rilevanti in caso di inadempienze gravi. E', tuttavia, opportuno ricordare che il riconoscimento di questa appartenenza è tutt'altro che immediato e scontato, oggi ancor più che nel passato, e che è assai improbabile che interventi punitivi possano, in un settore così diffuso e difficilmente verificabile, costituire condizioni in grado di produrre cambiamenti culturali significativi. Il materiale documentario - sovrabbondante, umile nella sua esteriorità, ingombrante e di difficile accesso - tende sempre più a essere considerato e trattato come prodotto di rapido consumo: il livello di attenzione e cura che pur si registra nella fase in cui i documenti sono attivi è quasi sempre seguito dai segni inequivocabili della trascuratezza non appena il trattamento dell'affare cui i documenti si riferiscono sia concluso: l'archivio corrente lascia il posto a un magazzino, l'organizzazione dei documenti si traduce nel tempo in un accumulo di materiali senza forma e senza strumenti di reperimento, il valore informativo scompare e rimane - nella sostanza, a fronte di benefici non percepibili nel breve e nel medio periodo - solo il costo del mantenimento di masse ingombranti di documenti scarsamente accessibili.

L'innovazione tecnologica ha introdotto rilevanti novità nei sistemi documentari, ma - per ora - non ha risolto le difficoltà della conservazione: ha accresciuto enormemente e positivamente l'interesse, e quindi gli investimenti, nella formazione e gestione dei documenti (aspetti peraltro assai rilevanti anche ai fini della conservazione futura), come testimoniano la stessa normativa sulla firma digitale e sul protocollo informatico e gli studi di fattibilità in materia di gestione informatica dei documenti predisposti da quasi tutte le grandi amministrazioni pubbliche italiane nell'ultimo biennio.

Nello stesso tempo, l'IT è destinata a produrre cambiamenti radicali nella forma stessa dei documenti e nei modi concreti della loro organizzazione e accumulazione in archivi, con conseguenze molto rilevanti proprio per quelle fasi del trattamento che seguono la gestione attiva, per quelle fasi cioè che da decenni sono gravemente trascurate dagli amministratori e lasciate alla cura dei soli archivisti di Stato o dei pochissimi archivisti qualificati presenti negli organici delle pp.aa. Gli specialisti del settore sono, certamente, preoccupati e consapevoli dei rischi di conservazione futura, ma non hanno le risorse, né tanto meno l'autorità necessarie per contrastare la crescente disattenzione in questo campo. Il seminario costituisce un'importante eccezione e, speriamo, un'inversione di tendenza almeno per quanto riguarda gli interlocutori istituzionali che hanno responsabilità e capacità di influire sui processi di informatizzazione del settore pubblico.

Rispetto ai documenti cartacei le prospettive per il mantenimento nel tempo, soprattutto nel lungo periodo, delle memorie digitali sono, quindi, assai più incerte, come già è stato ricordato in precedenza. Per di più, si riduce e si allontana la percezione stessa dei documenti informatici come patrimonio culturale da salvaguardare per l'eternità. Lo sviluppo tecnologico, infatti, smaterializzando i documenti e gli archivi, ne fa perdere di vista la corposità, complica la riconoscibilità del bene in quanto patrimonio accumulato nel tempo di sapere e testimonianza. Se la smaterializzazione da un lato rende possibile lo sviluppo della funzione informativa dei documenti, il valore di utilizzo nella vita pratica delle persone e degli enti (il binomio trasparenza/efficienza trova nell'automazione del sistema documentario uno strumento di crescita esponenziale), dall'altro rende meno evidente fino ad oscurare la funzione che i documenti svolgono nel lungo periodo in quanto memoria storica.

Non sono molti gli amministratori disposti a riconoscere oggi agli archivi digitali valore storico e culturale: gli enti spesso neanche si accorgono di possederli in quanto archivi (eppure sono ormai migliaia, solo nel settore pubblico, i sistemi informatici che hanno sostituito la tradizionale produzione documentaria, come ricorda un censimento predisposto dall'Aipa ormai alcuni anni fa e recentemente pubblicato sulla rivista "Archivi per la storia"^[1]). Gli stessi funzionari amministrativi del Ministero per i beni e le attività culturali - che pur dovrebbero essere sensibili al problema - reagiscono spesso con incredulità e disinteresse alla richiesta dei tecnici di settore, gli archivisti, di intervenire nei processi di normazione in corso in materia di documenti informatici per introdurre elementi significativi di salvaguardia dai rischi della dispersione futura.

Per parte loro, gli archivisti, pur riconoscendo l'urgenza del problema, ne temono anche la ripetitività nelle attività di routine, come ricorda Bruce Ambacher in un saggio dedicato alla ricostruzione del lavoro svolto a partire dal 1968 dal National Archives di Washington per l'acquisizione archivistica di documenti elettronici. La verifica e la validazione dei documenti elettronici, scrive Ambacher, "is time-consuming, labor intensive, boring and prone to error", oltre a richiedere una quantità notevole di ore di lavoro per un numero assai limitato di documenti esaminati^[2].

L'innovazione tecnologica consente soluzioni progressivamente più raffinate e differenziate per gli utenti e, quindi, sempre meno standardizzabili e riducibili a formati unitari per chi è impegnato nella conservazione. La cosiddetta "granularità" dell'informazione - un termine ricorrente nella letteratura tecnica per indicare la progressiva riduzione di oggetti un tempo unitari a un insieme di elementi informativi sempre più piccoli e tendenzialmente indipendenti - presenta rischi e difficoltà crescenti per chi ha il compito di rendere disponibile l'informazione all'interno di contesti storicamente definiti. Per fortuna, la fase attuale è caratterizzata dalla larga condivisione - su scala internazionale - dei fenomeni, che stanno producendo la perdita di stabilità fisica della memoria documentaria e allo stesso tempo la sua generale diffusione in ambienti non controllati:

- l'uso capillare delle tecnologie informatiche nell'attività amministrativa ma anche nella vita degli individui (lo scambio di messaggi di posta elettronica e la pubblicazione in rete hanno ad esempio trasformato radicalmente i modi tradizionali di comunicazione scientifica oltre che interpersonale, sostituendo non solo lo scambio verbale, ma anche qualunque altra forma di corrispondenza e di trasmissione di informazioni, nonché la diffusione stessa delle conoscenze);
- la diversificazione dei prodotti documentari e la loro virtualizzazione (lo stesso documento è spesso trattato, visto e conservato in versioni diverse nel medesimo contesto di produzione);

- l'allargamento dei confini del sistema documentario non più limitato allo spazio "amministrativo" dell'ente ma integrato nell'ambiente web, aperto alle interazioni con utenti sconosciuti e, perciò, oggetto di continue modifiche difficilmente identificabili e soprattutto controllabili.

Questi sviluppi implicano che la conservazione non sia neppure concepibile, se non si avviano le attività che la rendano possibile al momento stesso della formazione del sistema documentario informatico.

I problemi che si affollano sul tavolo dei conservatori del patrimonio documentario sono molteplici, sono tutti costosi, richiedono investimenti molto seri sul piano della ricerca, della formazione qualificata del personale, degli strumenti tecnologici e rischiano di avere un effetto paralizzante^[3]. Richiedono soluzioni organizzative, ma soprattutto implicano - per essere affrontati con qualche margine di speranza - la diffusione di standard e di metodi che siano il risultato di un serio lavoro di ricerca e di soluzioni applicative scalabili nella progettazione e nel mantenimento. Un primo obiettivo di questa riflessione è quello di gettare uno sguardo più attento - e quindi tecnico - alle ricerche in corso nella speranza di trovare una possibile continuità, un terreno già arato su cui fondare l'attività di conservazione che, come si è detto, riguarda innanzi tutto chi produce i documenti nell'esercizio delle sue funzioni. I soggetti produttori sono, infatti, in grado oggi - e si stanno avviando su questa strada - di formare e acquisire un'enorme quantità di materiale documentario in forma elettronica, ma non sembrano, per ora, altrettanto capaci di gestirlo nel tempo, soprattutto nel tempo lunghissimo della conservazione archivistica. Per avere successo in questo settore, servono, infatti, professionalità specifiche capaci di basarsi su strategie e standard tecnici da un lato e su modelli organizzativi consolidati dall'altro.

In questa sede, più che approfondire le questioni relative a scelte di natura organizzativa (peraltro assai rilevanti e impegnative in termini di risorse e investimenti), sembra più appropriato e, forse, urgente affrontare il tema degli standard, sia perché la definizione di regole tecniche comuni è, in materia documentaria, una importante occasione di riflessione teorica e merita senz'altro un'analisi dettagliata, sia per l'impatto che l'adozione o meno di norme condivise potrà avere sul problema specifico della conservazione a lungo termine di memorie digitali, a partire dalla possibilità stessa di immaginare un futuro duraturo. Il campo di indagine che richiede in questa fase e in questo settore una maggiore attenzione è quello tecnico-archivistico, che soffre dei ritardi accumulati e della difficoltà di riunire intorno a uno stesso tavolo profili professionali assai diversificati e non sempre ugualmente interessati al necessario approfondimento scientifico delle questioni di base della conservazione:

- l'analisi e la definizione dei requisiti di autenticità che garantiscano l'identificazione e "l'integrità possibile" delle memorie digitali;
- la definizione delle componenti informative (cioè i metadati) che costituiscono e connettono gli oggetti digitali destinati alla conservazione (i singoli documenti, le unità archivistiche, le informazioni di contesto, ecc.);
- lo studio di metodologie per la conservazione permanente che si basino soprattutto, se non esclusivamente, sull'utilizzo di formati e standard per la produzione di documenti indipendenti da qualunque piattaforma tecnologica. A questo fine è indispensabile - è stato già sottolineato - che tutte le componenti di un sistema documentario siano correttamente e tempestivamente identificate in quanto entità e nelle loro molteplici interrelazioni e che la documentazione disponibile sia tale da consentire di mantenere e testimoniare l'affidabilità del sistema, la correttezza dei processi di

migrazione, le modalità d'uso dei documenti, e molto altro ancora.

I tre ambiti di ricerca che si cercherà di approfondire in questa sede sono strettamente correlati: senza una chiara definizione dei requisiti di autenticità non è possibile identificare e valutare i metadati e la loro articolazione organizzativa. Allo stesso tempo, l'individuazione dei metodi di conservazione, l'adozione di standard nella fase stessa di formazione dei documenti, il loro utilizzo nei software applicativi implicano una approfondita e, soprattutto, coerente elaborazione di requisiti funzionali di autenticità e di informazioni logiche strutturate che descrivano le unità documentarie e archivistiche, i modi, i tempi, le responsabilità per la loro formazione o acquisizione, le modalità di accesso e uso, i processi di migrazione ed eventualmente i cambiamenti subiti nella fase ancora attiva e semi-attiva. In ognuno di questi settori nessun risultato è ancora certo, ma attività anche molto impegnative sono in corso e qualche primo risultato interessante comincia ad essere raggiunto, ad esempio per quanto riguarda l'identificazione di un modello logico per l'autenticazione dei documenti (presentato da Bill Underwood), l'analisi sui metadati archivistici avviata nell'ambito del forum organizzato dai colleghi olandesi, il progetto EAD per la descrizione dei documenti informatici, su cui interviene Ann Gilliland-Swetland e gli approfondimenti condotti sull'utilizzo di XML per la comunicazione e la documentazione amministrative affrontate in ambito internazionale da Zachary Coffin e per la ricerca nazionale da Daniele Tatti. All'origine ci sono comunque alcune importanti esigenze che non possono essere disattese se si intendono raggiungere soluzioni convincenti: chiarezza dei principi e della metodologia, identificazione degli schemi concettuali per il trattamento delle fonti digitali e individuazione dei metodi per la conservazione.

I requisiti di autenticità: la ricerca InterPARES

Un aspetto che merita subito un approfondimento è quello dei requisiti di autenticità, cioè la definizione di un quadro teorico in materia di archivi digitali. Dopo una fase di dibattito vivace che ha coinvolto scuole diverse e ha stimolato la riflessione di molti, soprattutto nel mondo anglosassone, il panorama degli studi condotti negli ultimi anni si è da un lato semplificato per quanto riguarda iniziative di ricerca di peso internazionale, dall'altro ha mostrato uno sconcertante livello di frammentazione e di ridondanza per la molteplicità di iniziative di dimensione locale^[4], che non sembrano per ora aver contribuito in modo significativo né alla riflessione teorica né alla predisposizione di strumenti operativi esportabili. Non poteva essere diversamente poiché la questione è molto impegnativa e può trovare risposte esclusivamente in un lavoro di indagine che coinvolga settori disciplinari diversi e risorse sufficienti a sostenere un tale lavoro di ricerca, come è stato ricordato da Reagan Moore dell'Università di S. Diego. Non è, perciò, un caso che le uniche due iniziative di ricerca in questo ambito (reciprocamente collegate e cooperative) siano quelle che hanno trovato il significativo sostegno di importanti istituzioni scientifiche nordamericane, più di altre impegnate sul fronte della conservazione permanente di documenti elettronici, il progetto internazionale InterPARES^[5], condotto dalla scuola di archivistica dell'Università del British Columbia e il progetto NPACI-NARA, sostenuto dal National Archives di Washington e dall'Università della California^[6].

Una specifica riflessione merita il progetto canadese InterPARES, cui l'Italia partecipa con un proprio team di ricercatori e di istituzioni. L'indagine costituisce - come ricordato da Ken Thibodeau - la base teorica e metodologica e il punto di partenza per lo sviluppo delle sperimentazioni future. Nasce dalla convinzione che una soluzione definitiva e complessiva al problema della conservazione documentale necessiti di un

impegno globale della comunità internazionale e di una seria ed efficace cooperazione tra discipline e ambiti professionali diversi al fine di pervenire alla comune definizione di:

- requisiti di autenticità dei documenti informatici, che consentano la verifica dei documenti dal punto di vista della loro integrità nel lungo periodo,
- strategie di valutazione e selezione dei documenti informatici, che individuino tempi e modi del trasferimento di responsabilità per la conservazione permanente,
- principi e procedure di autenticazione dei documenti elettronici nelle attività di conversione, copiatura, migrazione,
- criteri di descrizione coerenti con la natura archivistica del materiale trattato, con le esigenze della ricerca scientifica e con i bisogni di accesso di un'utenza non specialistica.

La ricerca è, come si è detto, in pieno svolgimento (si concluderà nel febbraio 2002), ma ha già consentito

- di elaborare uno schema - in fase di validazione mediante attività di ricognizione condotte su sistemi elettronici diversi - delle componenti logiche che formano la struttura dei documenti informatici allo scopo di identificare quegli elementi che siano necessari e sufficienti per assicurare l'autenticità dei documenti;
- di predisporre la base di discussione per la definizione dei requisiti da rispettare nelle fasi di verifica dell'autenticità (ad esempio in occasione del versamento dei documenti nelle istituzioni di conservazione^[7], ma anche nel corso degli interventi di migrazione);
- di sviluppare un modello formale della funzione conservativa dal punto di vista delle istituzioni cui sia affidata tale responsabilità.

Sin dalle prime fasi del lavoro è emerso che molte delle attività tradizionali di conservazione richiedono una riflessione teorica e metodologica più rigorosa se l'ambiente di produzione documentaria è digitale. Le ragioni sono numerose:

- la rilevanza degli investimenti necessari,
- la giusta e necessaria ambizione di rendere automatiche la gran parte delle attività di selezione, tenuta e descrizione dei documenti (al fine di rendere fattibile in termini di risorse umane e finanziarie la produzione, conservazione e utilizzo del patrimonio documentario contemporaneo in forma elettronica),
- la necessità di disegnare modelli onnicomprensivi e condivisibili.

Sebbene i risultati siano ancora parziali, alcune indicazioni meritano di essere ricordate proprio in relazione ai presupposti teorici della conservazione. Si tratta in particolare di quelle indicazioni che stabiliscono i requisiti per l'autenticità dei documenti, ovvero i requisiti che ne permettano l'**identificazione** (il mantenimento cioè dell'articolazione concettuale interna e del contesto di relazioni amministrative, giuridiche e documentarie) e di verificarne l'**integrità** (compatibilmente con i cambiamenti inevitabili determinati dai problemi di obsolescenza).

Innanzitutto, in ambiente elettronico, a causa proprio dell'obsolescenza tecnologica e della necessità di migrazione dei documenti, la conservazione a lungo termine può assicurare esclusivamente la **produzione di copie autentiche di documenti elettronici autentici** (l'integrità degli oggetti digitali è, inevitabilmente, legata alla necessità di prevedere modifiche anche significative nel flusso di bit che

costituisce il documento e le sue relazioni). Il problema dell'autenticità è, quindi, molto complesso e richiede una definizione su più piani:

- la assicurazione di autenticità per i documenti che sono ancora **attivi** presso il soggetto produttore,
- il mantenimento dell'autenticità per i documenti già **versati** negli archivi storici e destinati alla conservazione permanente,
- la verifica dell'autenticità per i documenti **in fase di trasferimento** dall'ambiente di produzione originario a quello di consultazione a fini di ricerca.

Quest'ultima costituisce senza dubbio un'attività cruciale e impegnativa: è in questa fase, infatti, che maggiori sono i rischi di perdite (anche accidentali) e di manomissioni, venendo meno la garanzia che deriva dal bisogno stesso del soggetto produttore di disporre di archivi affidabili e integri per svolgere le proprie attività quotidiane.

D'altra parte la verifica dell'autenticità, soprattutto nel caso di sistemi informatici, non può basarsi che sulla pre-esistenza nell'archivio corrente di condizioni e procedure che assicurino l'affidabilità del sistema e che a loro volta costituiscano gli strumenti per garantire e verificare l'integrità della produzione documentaria.

L'identificabilità dei documenti richiede, quindi, una serie di informazioni che devono essere mantenute a tempo indeterminato insieme ai documenti medesimi (nel profilo del documento), entrambi in forma leggibile e intelligibile:

- i dati di provenienza (organizzazione responsabile, autore),
- le componenti logiche interne (la cui quantità e qualità varia in base al tipo di documento, alla sua funzione e alla sua specifica forma),
- la registrazione univoca e con data certa,
- le relazioni documentarie che identificano le modalità di accumulazione, formazione e organizzazione stabile dell'archivio (classificazione e fascicolazione), la cui specifica natura varia, naturalmente, in base alla tipologia dei sistemi elettronici nel cui ambito i documenti si producono (database, sistemi di document management, sistemi interattivi, ecc.),
- l'impronta e il certificato relativi all'utilizzo della firma digitale quale elemento per l'autenticazione del documento al momento della sua formazione.

A proposito della necessità di disporre di principi solidi in materia di conservazione, è opportuno sottolineare, proprio a proposito della firma digitale - ma un discorso simile potrebbe essere fatto anche in riferimento all'uso di supporti specifici, ad esempio dei dischi ottici per l'archiviazione dei documenti - che si tratta di strumenti che contribuiscono a dare affidabilità al sistema e che, quindi, sono essenziali alla verificare dell'integrità delle entità documentarie, ma non possono sostituirsi alla stratificata serie di elementi e procedure che complessivamente determinano la natura autentica di un archivio. La firma digitale, in particolare, è solo un mezzo (importante, ma anche costoso e impegnativo) di autenticazione del documento nel momento stesso in cui esso si forma o si riproduce, ma certamente non è condizione sufficiente a garantire l'autenticità nel lungo periodo, innanzi tutto perché la coppia di chiavi e gli stessi registri dei certificati hanno durata limitata nel tempo (rispettivamente tre e dieci anni), ma anche perché la complessità dei sistemi documentari richiede la gestione e il mantenimento di un insieme articolato di informazioni contestuali di riferimento, che per la gran parte si formano nella fase attiva della gestione e che rischiano di perdersi nei trasferimenti

successivi, a meno che il sistema di trattamento dei metadati non ne preveda la cattura automatica secondo procedure prestabilite.

La verifica dell'integrità implica un'ulteriore serie di strumenti e sistemi di controllo che consentano di verificare tutte le azioni che hanno modificato i documenti, inclusi gli interventi di migrazione e di selezione:

- audit trail per l'accesso e per la trasmissione dei documenti,
- sistemi di tracciamento e di gestione sempre aggiornata delle informazioni di localizzazione,
- gestione controllata delle procedure di backup,
- ambienti di conservazione controllati ecc.

E' evidente che il tema dell'autenticità richiede di essere ulteriormente approfondito e costituisce un aspetto vitale dello stesso processo di informatizzazione. Si può, comunque, sostenere sin d'ora che è necessario :

- verificare che nessuna componente essenziale del documento, delle unità archivistiche e delle relazioni di contesto sia fatta oggetto di modifica,
- identificare, descrivere e documentare qualunque altro cambiamento.

I metodi per la conservazione

Perché si possa parlare di "conservazione archivistica" le componenti essenziali sono, quindi, molteplici e strutturalmente articolate. Più che salvare il flusso di bit che definisce un documento -, indispensabile soprattutto conservare le informazioni che rendono esplicita la sua rappresentazione e i suoi legami nel sistema documentario^[8].

In relazione ad altre aree di applicazione dell'informatica, il mondo degli archivi si caratterizza per la complessa e stratificata articolazione della produzione documentaria^[9], la cui peculiare natura originaria deve essere rigidamente salvaguardata per garantire la possibilità stessa della ricerca futura.

Un elemento vincolante è, naturalmente, quello del contenimento dei costi e della scalabilità delle soluzioni, tenuto conto della esiguità delle risorse finanziarie che sono normalmente a disposizione delle istituzioni archivistiche cui è affidato il compito della conservazione permanente delle memorie documentarie, incluse quelle digitali che le amministrazioni pubbliche e il settore privato hanno già cominciato a produrre in quantità rilevante. E' evidente che le possibilità di riuso sono legate a uno sviluppo significativo di standard e di metodi sperimentati che dovrebbero determinare un'effettiva diminuzione delle risorse impiegate e dei rischi di perdite (in particolare per quanto riguarda la conversione/migrazione delle applicazioni e la duplicazione delle informazioni).

Per quanto riguarda, poi, la scelta di metodi sperimentati e l'individuazione di standard per organizzare e gestire concretamente la funzione conservativa, l'incertezza è ancora notevole. Le soluzioni suggerite dagli esperti sono tutt'altro che consolidate, generalmente molto costose e, per ora spesso, prive di verifiche sul campo sufficienti. Si orientano sempre meno verso la conservazione delle tecnologie hardware e software (ormai considerata da alcuni anni una soluzione addirittura "utopistica")^[10], più frequentemente sostengono l'opportunità di sviluppare programmi di emulazione delle piattaforme tecnologiche originali, anche se si tratta di interventi che richiedono risorse elevate e non eliminano le rischiose e impegnative attività di migrazione né riducono le difficoltà dell'utenza costretta a misurarsi con strumenti assai diversificati

e spesso obsoleti anche dal punto di vista della presentazione e delle modalità di ricerca. La maggioranza degli esperti considera perciò tali ipotesi insufficienti e ribadisce l'urgenza di elaborare alternative fattibili ed efficaci basate soprattutto sul mantenimento dei dati relativi al contesto di produzione^[11]. Tra le proposte che hanno finora ottenuto i consensi maggiori e promettono sviluppi interessanti e utilizzabili in contesti operativi diversificati anche di piccole dimensioni, la conservazione in formati indipendenti dalle tecnologie - basati sull'uso di linguaggi di marcatura (SGML/XML) - della rappresentazione originaria dei documenti e dei metadati di contesto e di relazione sembra destinata - nel medio e lungo periodo - a significativi sviluppi. Si tratta di una soluzione che risponde ai requisiti di base dell'Open System Interconnection, poiché utilizza standard non proprietari, applicabili a sistemi informatici diversi, largamente diffusi, comprensibili anche da parte di non specialisti, indipendenti dall'hardware e ben documentati e, oggi, anche in via di implementazione^[12].

XML offre un metodo diffuso, a basso costo e scalabile per affrontare la diversificazione e la frammentazione della produzione documentaria e delle sue articolazioni, la sua ricchezza informativa e il peso, finora insostenibile per i bilanci limitati degli enti culturali, delle innovazioni tecnologiche.

Lo standard apre ulteriori e rilevanti possibilità per lo sviluppo di sistemi documentari informatici, soprattutto perché consente, oltre alla gestione dei riferimenti esterni al documento e alle sue partizioni, anche il trattamento della struttura logica e semantica dei contenuti. Questi sviluppi si possono tradurre nella decisione di:

- promuovere, all'interno di un'organizzazione, interventi di razionalizzazione e semplificazione delle tipologie documentarie mediante la definizione di rappresentazioni specifiche con lo scopo di ottimizzare l'elaborazione automatica dei documenti, garantire coerenza, qualità e uniformità dei materiali^[13],
- sviluppare strumenti di recupero e riutilizzo di documenti (o di componenti interne) ai fini di una distribuzione/condivisione di contenuti destinati a durare nel tempo,
- gestire formati multipli,
- utilizzare i sistemi di validazione XML anche a fini di sicurezza e di integrità,
- controllare e ottimizzare i cicli di gestione dei documenti.

E', tuttavia, importante sottolineare che XML può svolgere una funzione significativa nei processi di automazione del settore documentario, anche in termini ricordati di contenimento dei costi ed efficienza dei risultati, se è accompagnato da un uso diffuso di DTD. Le Document Type Definition - è stato recentemente ricordato da Charles Goldfarb e Paul Prescod^[14] - migliorano, infatti "la permanenza, la longevità e l'ampio riutilizzo dei propri dati, insieme alla prevedibilità e all'affidabilità della loro elaborazione". Tuttavia, lo sviluppo di DTD è una questione che rimette al centro della progettazione i problemi di struttura logica e concettuale e, naturalmente, richiede un approccio seriamente interdisciplinare e soprattutto presuppone un'effettiva volontà di cooperazione per la definizione di regole comuni, se non di veri e propri standard di settore. E', ad esempio, il caso dell'Encoded Archival Description promosso dalla Library of Congress^[15] (Ann Gilliland Swetland) oppure della circolare in corso di elaborazione da parte dell'Autorità per l'informatica per la definizione di una DTD per lo scambio di documenti in rete e lo sviluppo di sistemi avanzati di interoperabilità tra pubbliche amministrazioni (Marcello Tatti)^[16].

Anche a fronte di qualche prospettiva incoraggiante che vede finalmente la luce, è bene, tuttavia, non farsi illusioni sulle difficoltà che comunque attendono archivisti, specialisti di IT e amministratori impegnati nel campo conservazione: il monitoraggio delle tecnologie e l'attenzione continua ai costi per la migrazione sono, ad esempio, solo due attività, peraltro assai complesse, all'interno di un complesso di iniziative e procedure che non sono state ancora identificate con chiarezza, per le quali non esiste ancora sufficiente preparazione e che comunque non potranno essere in futuro né ignorate né sottovalutate nella stessa progettazione e nella realizzazione di sistemi documentali interamente informatici.

Il rischio maggiore, sottolineato anche da una collega bibliotecaria del Regno Unito, è quello di restare affascinati e allo stesso tempo paralizzati da tutte le opportunità e dai corrispondenti obblighi^[17]. E' quanto sembra accadere, talvolta, alle amministrazioni italiane e alle stesse istituzioni che hanno per mandato il compito della conservazione a lungo termine. E' vero che gli obiettivi sono difficili da affrontare, le conoscenze insufficienti e le risorse disponibili addirittura ridicole. D'altra parte l'assenza di iniziativa si traduce in una negligenza che comporta la perdita completa dei nuovi archivi, un esito che nessuna comunità democratica può permettersi.

[1] Carlo Batini, Guglielmo Longobardi, Isabella Siciliani, *Repository delle basi di dati della pubblica amministrazione*. Studio Aipa, in "Archivi per la storia", 1999, n. 1-2, pp. 377-441.

[2] Bruce Ambacher, *The evolution of processing procedures for electronic records*, rapporto presentato all'Annual Meeting della Society of American Archivists, Denver, Colorado, 29 agosto 2000.

[3] Nancy E. Elkington, *Heritage, scholarship and preservation in the digital age*, in *Making information available in digital format. Perspectives from Practitioners*, St Crispins, Norwich 1999, p. 40.

[4] E' significativo l'esito della ricognizione condotta dall'Unione europea sull'esistenza di linee guida per la conservazione digitale in tutto il settore dei beni culturali: dopo un anno di lavoro e numerose interviste, ricognizioni e analisi, il gruppo di lavoro ha stabilito che non esistevano al 1998 linee guida in grado di affrontare il problema complessivo della conservazione digitale nel settore culturale e che "long-term perspectives on preserving access to digital archives still require fundamental work". Cfr. Marc Fresko, Kenneth Tombs, *Digital preservation guidelines: the state of the art in libraries, museums and archives*, Brussels, European Commission, DG XIII/E, 1998.

[5] L'indagine costituisce la prosecuzione del lavoro svolto nel corso del precedente e già ricordato progetto sulla formazione e gestione dei documenti attivi e affronta il problema specifico della conservazione a lungo termine dell'integrità e autenticità dei documenti elettronici. Alla ricerca partecipano undici Paesi (Australia, Canada, Cina, Francia, Irlanda, Italia, Olanda, Portogallo, Stati Uniti, Svezia, UK) e un team di industrie farmaceutiche. I materiali del progetto (che si concluderà nel febbraio 2002) sono disponibili nel sito dedicato alla diffusione dei risultati della ricerca <http://www.interpares.org>. Alcuni documenti sono stati recentemente pubblicati sulla rivista "Archivi per la storia", 1999, n. 2.

[6] Il progetto è analizzato nell'intervento di Reagan Moore e di Kenneth Thibodeau. Materiali della ricerca sono disponibili al seguente indirizzo: <http://www.sdsc.edu/NARA/Publications/collections.html>. Nell'ambito di tale indagine, si è recentemente avviata ufficialmente una collaborazione specifica (finanziata a partire dal 2001 dal Ministero per l'università e la ricerca scientifica, tra i promotori e sostenitori della ricerca e l'Università di Urbino (a sua volta capofila di un nucleo di istituzioni e centri di ricerca italiani) sull'uso di XML per la definizione e strutturazione dei metadati necessari alla conservazione di documenti elettronici autentici.

[7] Si vedano in particolare i materiali prodotti dalla Authenticity Task Force del progetto: *Research methodology statement*, *Template for analysis* e *Case Study protocol and questionnaire*, in "Archivi per la storia", 1999, n. 1-2, pp. 263-337.

[8] Tutti gli autori sottolineano la difficoltà di entrambi questi obiettivi: da un lato la necessità della migrazione, che si impone inevitabilmente e ripetutamente nell'attività di conservazione, può introdurre cambiamenti anche significativi nel flusso di bit, dall'altro la sempre più diffusa rappresentazione ad oggetti tende a rendere trasparenti agli utenti le informazioni "di contesto", che invece devono essere identificate e trattate in modo esplicito. Si veda in proposito Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System* (OAIS), cit., sezioni 3.3 e 3.4.

[9] Per un'analisi dei sistemi documentari informatici e dei requisiti funzionali che garantiscono una corretta formazione e gestione, cfr. Maria Guercio, *La formazione dei documenti in ambiente digitale*, in *Gli archivi del futuro. Il futuro degli archivi*. Cagliari, 29-31 ottobre 1998 (numero monografico di "Archivi per la storia", 1999, n. 1-2), pp. 21-58.

[10] J. Rothenberg, *Ensuring the longevity of digital documents*, in "Scientific American", 1995, 1, pp. 24-29; S. Ross-A. Gow, *Digital archaeology: rescuing neglected or damaged data resources*, London 1999.

[11] Cfr il citato rapporto di Marc Fresko, Kenneth Tombs, *Digital preservation guidelines* e il documento proposto come standard ISO dal Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System* (OAIS).

[12] C.F.Cargill, *information Technology Standardization. Theory, Process and Organizations*, Bedford 1989.

[13] L'utilizzo di XML accentua l'importanza della struttura semantica dei documenti anche perché lo standard consente di distinguere in modo chiaro tra elementi e attributi, cioè - in termini di analisi del documento archivistico - tra i dati che costituiscono la struttura costitutiva generale per tipi di documento e gli attributi intesi come informazioni di secondo livello (proprietà degli oggetti e non sue parti). Cfr in proposito Charles F. Goldfarb e Paul Prescod, *XML*, Milano, McGraw-Hill Italia, 1999, p. 400.

[14] *Ibidem*

[15] L'Encoded Archival Description (EAD) è il risultato di un progetto di ricerca avviato nel 1993 dall'Università di Berkeley sull'uso dei linguaggi di marcatura (allora SGML) per la pubblicazione di strumenti di ricerca in ambiente digitale. Cfr Giovanni Michetti, // *linguaggio SGML per la descrizione archivistica: analisi di un caso*, in "Archivi e computer", 2000, 1, pp.7-33.

[16] Il provvedimento, approvato con dpcm 31 ottobre 2000 e disponibile all'indirizzo web dell'Aipa ([http://www.aipa.it/attivita\[2/protocollo13/gestflus\[2/gedoc2\[2/index.asp](http://www.aipa.it/attivita[2/protocollo13/gestflus[2/gedoc2[2/index.asp)) conclude, insieme alla delibera Aipa n. 51 del 23 novembre 2000 sulla formazione, gestione e conservazione dei documenti prodotti dalle pubbliche amministrazioni, la lunga serie di disposizioni concernenti la gestione informatica dei documenti amministrativi, avviata con il dpr 428/1998 sulla tenuta del protocollo informatico. Le regole tecniche, applicative del dpr citato, in corso di approvazione da parte della Presidenza del Consiglio dei ministri, dedica una intera sezione alle modalità di trasmissione e registrazione dei documenti informatici e introduce l'obbligo per lo scambio dei dati relativi alla segnatura di protocollo dell'utilizzo dello standard XML e delle DTD elaborate dal Centro tecnico per la rete unitaria della p.a. In particolare, l'articolo 19 stabilisce le informazioni da includere nella segnatura: oggetto, mittente e destinatario costituiscono le informazioni obbligatorie, cui si possono aggiungere i dati relativi alla persona o all'ufficio all'interno della struttura destinataria cui si presume sia affidato il trattamento del documento, l'indice di classificazione, l'identificazione degli allegati, il procedimento e il suo trattamento e tutte le informazioni che le amministrazioni specifiche vorranno concordare nell'ambito di rapporti reciproci.

[17] Nancy E. Elkington, *Heritage, scholarship and preservation in the digital age...cit.*, p. 40.