



Autorità per l'informatica nella Pubblica Amministrazione

La conservazione dei documenti
informatici - Aspetti organizzativi e
tecnici

Seminario di studio
Roma 30 ottobre 2000

Programma

I SESSIONE: LA FUNZIONE CONSERVATIVA. PROBLEMI, STRATEGIE, ESPERIENZE,

- 9:00 Apertura del seminario**
Salvatore Italia Direttore Ufficio Centrale per i beni archivistici
Guido Mario Rey –Presidente dell'Autorità
- 9:15 Relazione introduttiva**
Carlo Batini (Autorità per l'Informatica nella Pubblica Amministrazione)
- 9:35 La gestione e la conservazione dei documenti informatici nelle pubbliche amministrazioni italiane (The management and the preservation of electronic records in the Italian public administrations)**
Antonio Massari (Autorità per l'Informatica),
- 10:10 Information Management Architecture for Persistent Object Preservation**
Reagan W. Moore (University of California)
- 10:40 Intervallo**
- 11:00 The strategies for the preservation. Projects and perspectives in the US Federal Government**
Ken Thibodeau (National Archives USA)
- 11:30 The Swedish programs for the electronic records preservation and access**
Kristiansson Gorlan (Riksarkivet, Sweden)
- 12:00 Responding to the challenges and opportunities of ITC: the new record manager**
Seamus Ross (University of Glasgow)
- 12:30 Discussione**
- 13:00 Intervallo per il pranzo**

II SESSIONE. I METODI PER LA CONSERVAZIONE, IL RUOLO DEI METADATI, LE POTENZIALITÀ DI XML

- 14:00 Relazione introduttiva**
Maria Guercio (Università di Urbino)
- 14:30 A logical model for the electronic records authentication**
Bill Underwood (Georgia Tech Institute)
- 15:00 Metadata: Archival Concept or IT Domain?**
Peter Horsman (IT Committee, International Council on Archives, president)
- 15:30 The potential of markup languages to support descriptive access to electronic records: the EAD standard**

Anne Swetland Gilliland (University of Los Angeles)

16:00 XML standards for business-to-business and business to government communication

Zachary Coffin (eXtensible Business Reporting Language)

16:30 XML, uno standard per gli archivi informatici (XML, a standard for the electronic records)

Daniele Tatti (Autorità per l'informatica)

Indice

Indice 4

Premessa 6

THE EVOLUTION OF PROCESSING PROCEDURES FOR ELECTRONIC RECORDS 7

ARCHIVAL PRINCIPLES IMBUED IN ELECTRONIC RECORDS PROCESSING 7

ACCESSIONING 8

ARCHIVAL PRESERVATION SYSTEM 9

ARCHIVAL ELECTRONIC RECORDS INSPECTION AND CONTROL SYSTEM 10

ARCHIVAL MANAGEMENT INFORMATION SYSTEM 10

GAPS DATABASE 11

DOCUMENTATION 11

PRESERVATION 12

REASONS FOR PAST SUCCESS 13

EMERGING ISSUES 14

History of NARA's Electronic Records Program 15

An Historical Perspective on Appraisal of Electronic Records, 1968-1998 21

Record/ Non-Record, Not Valuable/Valuable 21

Applying Traditional Archival Theory of Appraisal 22

Applying Traditional Records Management Techniques 23

Innovation, trying new approaches 23

Knowledge-based Persistent Archives 26

Abstract 26

1. Introduction 26

2. Knowledge-based Archives 27

2.1 Archive Accessioning Process: 29

2.2 Archival Representation of Collections: 31

3. Relationships between NARA and other Agency projects: 32

Acknowledgements: 32

NARA and Electronic Records: A Chronology 34

The potential of markup languages to support descriptive access to electronic records: The EAD standard 36

Abstract: 36

Introduction 36

Describing Electronic Records 38

Encoded Archival Description 39

Using EAD to Describe Electronic Records 40

Conclusion 42

Preservation and migration of electronic records: the state of the issue 43

The problem of preserving electronic records 43

Approaches to the problem of preserving electronic records 43

The need for an archival approach to preserving electronic records 44

An architecture for archival preservation 45

- 1. the right data was put into storage properly; 45**
- 2. either nothing happened in storage to change this data or alternatively any changes in the data over time are insignificant; 45**
- 3. all the right data and only the right data was retrieved from storage; 45**
- 4. the retrieved data was subjected to an appropriate process, and 45**
- 5. the processing was executed correctly to output an authentic reproduction of the record. 45**

Collection-based persistent object preservation 46

Viability of the persistent object preservation method 47

Conclusion 47

Responding to the Challenges and Opportunities of ICT: The New Records Manager 49

XML per la conservazione dei sistemi documentari informatici 50

Premessa

Il seminario, organizzato dall'Autorità e dall'Ufficio Centrale per i Beni Archivistici del Ministero per i Beni e le Attività Culturali, ha la finalità di affrontare la questione cruciale della conservazione a lungo termine dei documenti informatici, da un lato sensibilizzando le amministrazioni pubbliche che hanno già avviato o programmato processi di automazione dei sistemi documentari, sia di avviare - anche attraverso un confronto internazionale - l'analisi e la verifica delle modalità e degli strumenti che consentono il mantenimento nel tempo dell'integrità e dell'autenticità delle memorie digitali nonché la loro accessibilità. Il tema ha acquistato attualità e rilevanza crescenti in relazione alla larga diffusione di tecnologie dell'informazione e della comunicazione per la gestione dei flussi amministrativi e documentari. Richiede, tuttavia, la soluzione di numerosi problemi di natura organizzativa e tecnica che gruppi internazionali di ricerca e istituzioni nazionali di tutto il mondo hanno cominciato a studiare da qualche anno con impegno crescente soprattutto allo scopo di individuare soluzioni scalabili e di basso costo che rendano possibile anche alle organizzazioni di piccole e medie dimensioni avviare interventi generali di automazione dei processi amministrativi e delle attività di trasmissione e tenuta dei documenti.

In particolare, **nella prima sessione**, il seminario si propone di offrire alle amministrazioni pubbliche italiane - soprattutto ai responsabili dei sistemi informativi e ai responsabili dei sistemi documentari - un panorama della specifica situazione nazionale e dei progetti e delle esperienze operative più interessanti condotte in altri Paesi (Stati Uniti e Svezia) in stretta collaborazione con più ampi gruppi di ricerca internazionali, insieme a un'analisi delle esigenze di formazione e riqualificazione dei profili professionali destinati ad assumere il nuovo difficile ruolo di conservatore delle memorie digitali.

La seconda sessione è destinata ad esaminare le questioni più strettamente tecniche della conservazione, in particolare i diversi metodi oggi disponibili per affrontare l'obsolescenza tecnologica e, in special modo, le potenzialità dei linguaggi di marcatura (SGML, XML) per il trattamento e la gestione dei metadati che garantiscono la tenuta a lungo termine del patrimonio documentario e lo sviluppo di standard specifici (Encoded Archival Description).

XML per la conservazione dei sistemi documentari informatici

Maria Guercio - Università degli Studi di Urbino

Perché XML per gli archivi?

Da alcuni anni c'è un grande interesse per i linguaggi di marcatura da parte degli amministratori e dei conservatori di beni culturali, in particolare degli archivisti e dei bibliotecari, che sembrano aver finalmente trovato uno standard adeguato alle complesse esigenze informative, di comunicazione e tenuta delle memorie documentali digitali.

Nel caso specifico degli archivi elettronici, nonostante gli sforzi compiuti da numerosi centri di ricerca istituzionali e accademici³¹, non si sono ancora raggiunti risultati convincenti sia nella definizione di strategie sia nella elaborazione di procedure e soluzioni informatiche, soprattutto per quanto riguarda i problemi della conservazione nel tempo. La funzione conservativa in ambiente digitale costituisce, del resto, uno dei compiti più difficili e impegnativi per una serie di ragioni, di cui la principale riguarda la natura contrastante e apparentemente inconciliabile dell'obiettivo medesimo: mantenere l'integrità certa dei documenti informatici e, al contempo, assicurare un'accessibilità³² che, a causa dell'obsolescenza tecnologica, implica continui interventi di copiatura, conversione, migrazione e, quindi, continue modifiche nella struttura dei bit che costituiscono il documento. Si tratta per gli archivisti di una vera e propria sfida che richiede in primo luogo la definizione di una base concettuale e metodologica e un quadro di riferimento teorico solidi, ma anche un grande sforzo organizzativo e tecnologico, risorse finanziarie elevate, tecnici di altissima professionalità e lo sviluppo di prodotti informatici in grado di garantire la produzione e il trattamento dei documenti con modalità di routine compatibili con l'esigenza della loro conservazione permanente.

Per quanto riguarda in particolare il 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 mostra uno sconcertante livello di frammentazione e di ridondanza per la molteplicità di iniziative di dimensione locale³³ che non sembrano per ora contribuire in modo significativo né alla riflessione teorica né alla predisposizione di strumenti operativi esportabili. D'altra parte la questione, come si è detto, è molto impegnativa e non può trovare risposte esclusivamente in un lavoro di indagine che coinvolga settori disciplinari diversi e risorse adeguate. Non è, perciò, un caso che le uniche due iniziative di ricerca oggi attive in questo ambito (reciprocamente collegate e cooperative) siano quelle che hanno trovato il sostegno delle istituzioni

³¹ Negli ultimi anni sono stati condotti numerosi studi sui documenti elettronici, che tuttavia si sono concentrati - con esiti interessanti - soprattutto sul problema della formazione di sistemi documentari informatici. Le indagini di maggior rilievo internazionale sono state, in particolare, quelle condotte dalla Università di Pittsburgh e dalla Università del British Columbia (Vancouver, Canada), entrambe concluse nel 1997. La ricerca canadese è stata realizzata d'intesa con il Dipartimento della difesa degli Stati Uniti che, a conclusione del lavoro svolto, ha elaborato le regole per la certificazione dei programmi di automazione della gestione documentaria destinati all'amministrazione federale degli Stati Uniti. Per maggiori informazioni si vedano i materiali disponibili ai seguenti indirizzi: <http://www.lis.pitt.edu/~nhprc/> per la ricerca dell'Università di Pittsburgh; <http://www.slais.ubc.ca/users/duranti/> per la ricerca della Università del British Columbia e <http://jirc.fhu.disa.mil/recmgt/> per lo standard definito dal Dipartimento della difesa degli Stati Uniti "Standard 5015.2 - Design Criteria Standard For Electronic Records Management Software Applications".

³² Ken Thidobea, Reagan Moore, Chaitanya Baru, *Persistent object preservation: Advanced computing infrastructure for digital preservation*, in *Proceedings of the DLM-Forum on electronic records. European citizens and electronic information: the memory of the Information Society*. Brussels, 18-19 October 1999, Luxembourg, Office for official publication of the European Communities, 2000, pp. 113-120.

³³ 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.

scientifiche nazionali nordamericane, il progetto internazionale InterPARES³⁴, 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 California³⁵.

Per quanto riguarda gli aspetti teorici e metodologici, il progetto canadese, cui l'Italia partecipa con un proprio team di ricercatori e di istituzioni, costituisce senz'altro l'iniziativa di ricerca più significativa. Il progetto 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:

- strategie di valutazione e selezione dei documenti informatici, che individuino tempi e modi del trasferimento di responsabilità per la conservazione permanente,
- standard e norme per i supporti,
- 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,
- standard e procedure per la tutela della privacy e in materia di copyright.

La ricerca è, come si è detto, in pieno svolgimento, ma ha già consentito di elaborare un primo schema - in fase di validazione mediante attività di ricognizione condotte su sistemi elettronici diversi - delle componenti logiche che formano la struttura dei documenti informatici³⁶.

Per quanto riguarda, invece, la scelta di metodi sperimentati per organizzare e gestire concretamente la funzione conservativa, l'incertezza è notevole. Le soluzioni suggerite dagli esperti sono tutt'altro che consolidate, generalmente molto costose e, per ora, prive di verifiche sul campo. Si orientano verso la conservazione delle tecnologie hardware e software oppure sostengono l'opportunità di sviluppare programmi di emulazione delle piattaforme tecnologiche originali. In entrambi i casi 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 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 efficaci³⁷. 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 periodo - a significativi sviluppi.

Anche in questo caso la scarsa letteratura disponibile in materia³⁸ non offre per il momento indicazioni univoche e convincenti sulla strada da seguire, limitandosi a individuare vantaggi e svantaggi di ognuna delle ipotesi formulate. Inoltre, la scarsità di risorse e l'insufficienza di esperienze e conoscenze finora accumulate dalle istituzioni competenti hanno costituito fino ad oggi ostacoli quasi insormontabili alla rapida individuazione di soluzioni condivisibili. Non è, quindi, fuori di luogo l'allarme

³⁴ 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 al sito <http://www.interpares.org>. Alcuni documenti sono stati recentemente pubblicati sulla rivista "Archivi per la storia", 1999, n. 2.

³⁵ Materiali della ricerca sono disponibili al seguente indirizzo:
<http://www.sdsc.edu/NARA/Publications/collections.html>.

³⁶ 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.

³⁷ 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).

³⁸ Una bibliografia accurata si trova in Marc Fresko, Kenneth Tombs, *Digital preservation guidelines...*cit. Si vedano anche le indicazioni contenute nella *Nota bibliografica sul documento elettronico. 1986-1998*, pubblicata nel citato numero di "Archivi per la storia" 1999, n. 1-2, pp.347-375.

degli archivisti per il futuro delle memorie digitali, anche se l'ultima delle soluzioni proposte, su cui questo contributo si sofferma in modo particolare, sembra rispondere a molte delle esigenze della conservazione permanente e offrire qualche speranza alle preoccupazioni più diffuse.

Per comprendere meglio la natura del problema e valutare le aspettative suscitate dal nuovo standard presso la comunità archivistica nazionale e internazionale, è tuttavia necessario identificare, sia pure per grandi linee, i nodi teorici e pratici legati alle attività di gestione, uso e conservazione "archivistica" dei sistemi documentari informatici³⁹. Gli oggetti che devono essere identificati e mantenuti perché si possa parlare di "conservazione archivistica" sono molteplici e strutturalmente articolati. Non basta, infatti, salvare il flusso di bit che definisce un documento, ma è indispensabile anche conservare le informazioni che rendono esplicita la sua rappresentazione e i suoi legami nel sistema documentario⁴⁰. Sono inoltre essenziali le modalità di rappresentazione e comunicazione che implicano l'adozione di parametri uniformi di descrizione e di accesso, tutt'altro che semplici da determinare in un settore che, non a caso, è arrivato molto tardi e con molte resistenze - rispetto ad altri analoghi campi disciplinari - all'accettazione di pratiche normalizzate. Assai differenziata è, peraltro, anche l'utenza dei sistemi archivistici per il grado di conoscenza del sistema documentario, per le modalità di interrogazione, per la natura stessa delle ricerche. In relazione ad altre aree di applicazione dell'informatica, il mondo degli archivi si caratterizza per la complessa e stratificata articolazione della produzione documentaria⁴¹, la cui peculiare natura originaria deve essere rigidamente salvaguardata per garantire la possibilità stessa della ricerca futura.

A chi progetta interventi di automazione in questo ambito la specificità di tale materiale costituisce allo stesso tempo un vincolo e un'opportunità: le potenzialità di continua trasformazione offerte dalle tecnologie dell'informazione sono tali per gli archivisti solo se riferite alla ricchezza informativa e alla facilità di recupero di contenuti strutturati in modo logico e relazioni significative. Diventano invece gravi rischi da eliminare o, quantomeno, controllare e limitare se l'obiettivo sia quello di garantire l'autenticità e l'integrità del sistema nel tempo. Per dare solo un'idea dell'articolazione del sistema documentario e della natura strutturata dei suoi contenuti e relazioni, si ricorda che:

- l'archivio non è un semplice insieme di documenti, ma un insieme complesso di entità a loro volta costituite di sottopartizioni di diversa tipologia (subfondo, serie, sottoserie, fascicolo, sottofascicolo, unità documentaria),
- ciascuna sottopartizione è identificata e descritta mediante informazioni di natura generale condivisibili anche in un contesto internazionale⁴² (segnatura archivistica, denominazione, estremi cronologici, consistenza, ecc.), integrata da eventuali ulteriori dati significativi,
- le stesse unità documentarie non sono riducibili a semplice informazione testuale, ma si strutturano in una serie di componenti riconosciute all'interno di uno schema generale⁴³ e facilmente identificabili (autore, destinatario, data, oggetto, testo, indicazione di allegati, ecc.),

³⁹ E' indispensabile tenere sempre distinte le attività di memorizzazione da quelle di conservazione, che implicano la decisione di mantenere il patrimonio nel lungo periodo e la stabilità dei documenti e delle relazioni sia archivistiche che amministrative, un sistema di accesso e di ricerca elaborato in modo adeguato alle esigenze di ricerca di una comunità ampia.

⁴⁰ 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.

⁴¹ 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.

⁴² Si fa qui riferimento allo standard per la descrizione archivistica, *International Standard for Archival Description* approvati dal Consiglio internazionale degli archivi nel 1996 e recentemente aggiornato dal Committee for archival description. Il testo è disponibile al seguente indirizzo: <http://www.anai.org>.

⁴³ La disciplina che studia il documento e le sue componenti fisiche e logiche è la diplomatica, che negli ultimi anni, per merito di alcuni archivisti italiani, ha allargato il suo campo di indagine tradizionale, limitato in passato all'età

- tali unità, a loro volta, possono essere organizzate per tipologie specifiche che condividono un maggior numero di dati significativi rispetto allo schema generale comune,
- le informazioni relative al contesto giuridico, amministrativo e organizzativo di un sistema archivistico sono rilevanti ai fini della intelligibilità stessa delle testimonianze documentarie. Tali informazioni di contesto (ufficio di assegnazione, ufficio di provenienza, tipo di procedimento/processo amministrativo, responsabile del procedimento) devono essere, perciò, catturate dal programma informatico e mantenute in via definitiva sia a fini giuridici che per la comprensione amministrativa e storica del materiale prodotto.

Uno dei nodi centrali per la tenuta e l'accesso nel tempo ai documenti d'archivio informatici è, quindi, quello di assicurare il mantenimento in modo stabile e in ambiente sicuro della molteplicità delle informazioni strutturate - di attributi e relazioni di metadati -, che abbiano reso funzionale il sistema documentario nella fase attiva e che, per tale ragione, saranno conservate e messe a disposizione degli utenti per le future attività di ricerca storica, scientifica, amministrativa.

Non è questa la sede per affrontare in modo approfondito la questione - assai dibattuta in sede teorica - del ruolo e del trattamento dei metadati archivistici. Il tema - su cui scuole di archivistica diverse hanno discusso e discutono senza trovare ancora un punto di vista comune - riguarda le modalità di trattamento di queste meta-informazioni, sia nella fase attiva del sistema documentario che nel corso del trasferimento dei documenti destinati alla conservazione permanente nelle istituzioni archivistiche competenti⁴⁴. Nessun autore nega il loro valore strategico in ambito archivistico, sia per la complessità di strutturazione dei sistemi documentari, sia per l'univocità degli elementi che li costituiscono. A differenza di altri materiali, ad esempio i beni librari, i documenti d'archivio e, ancor più, i riferimenti di natura contestuale variano in ogni struttura organizzata e per ciascuno specifico ordinamento. Anche nel caso in cui si conservino più esemplari del medesimo documento in diverse aree della stessa organizzazione, è quasi sempre necessario ridefinire e mantenere i dati di contesto che cambiano in base alle specifiche procedure di classificazione e ordinamento. Le singole componenti (i documenti), ma soprattutto le reciproche relazioni (*il vincolo archivistico*) hanno all'interno di un archivio valori propri che si traducono in un'autonoma, ben definita gestione di metadati.

La strada più semplice - ma non per questo necessariamente più adeguata - è quella sostenuta da alcuni ricercatori nordamericani (in particolare da David Bearman) e ripresa successivamente dalla scuola archivistica australiana, che ritiene possibile incapsulare tutti i metadati all'interno della singola entità documentaria⁴⁵. In realtà si tratta di una proposta che semplifica senza risolvere e, perciò, impoverisce il problema della conservazione dei documenti digitali, poiché appiattisce su due soli livelli la struttura informativa da mantenere: i documenti da un lato, i metadati identificativi del singolo documento (ad esempio, i dati del profilo documentario) dall'altro. In realtà le componenti informative che si creano nella vita di un archivio e che sono indispensabili per la sua sopravvivenza come bene culturale sono ben più ricche di articolazioni e richiedono procedure di gestione più attente alla stratificazione storica delle attività e dei dati, a prescindere dal metodo e dagli strumenti impiegati per il loro mantenimento.

La questione che deve essere affrontata è, quindi, quella di identificare le strutture e gli schemi logici di metadati corrispondenti agli oggetti informativi che si intendono salvaguardare a fini storici e alle attività e funzioni di sistema di cui è necessario tenere traccia storica nel lungo periodo (organigrammi del soggetto produttore di documenti, piani di classificazione e repertori dei fascicoli, sistemi di registrazione

medievale e moderna, fino a comprendere non solo i documenti contemporanei, ma anche i documenti elettronici. Cfr. Paola Carucci, *Il documento contemporaneo. Diplomatica e criteri di edizione*, Roma, Nuova Italia Scientifica, 1987 e, più recentemente, Luciana Duranti, *Diplomatics: new uses for an old science*, Washington, SAA, 1999.

⁴⁴ L'intervento sulle meta-informazioni deve essere "precocemente", riguardare cioè il sistema attivo, poiché la maggior parte delle informazioni che garantiscono la continuità dell'accesso sono disponibili esclusivamente nell'archivio corrente (ad esempio i dati relativi alla struttura amministrativa che produce i documenti, lo schema logico di un database, la documentazione di un programma applicativo). Interventi tardivi di recupero sono, talvolta, possibili, ma sono di gran lunga più costosi e impegnativi.

⁴⁵ David Bearman, Ken Sochats, *Metadata Specifications Derived from Functional Requirements: A Reference Model for Business Acceptable Communications*, documento disponibile al seguente indirizzo www.lis.pitt.edu/~nhprc.papers/model.html. Sulle conclusioni raggiunte dagli archivisti australiani in materia, cfr. Sue McKemmish, *Australian Research and Development Initiatives*, in "Archivi per la storia", 1999, n. 1-2, pp. 197-206.

e autenticazione, elenchi dei procedimenti/processi amministrativi, interventi di copiatura, conversione, migrazione, ecc.).

In conclusione, le informazioni di riferimento ai documenti destinate ad essere oggetto di trattamento conservativo sono nel caso degli archivi così numerose che richiedono di essere organizzate per componenti funzionali (metadati di contesto amministrativo, quali ad esempio quelli relativi alla struttura organizzativa del soggetto produttore del sistema documentario, metadati di contesto documentario, ad esempio le informazioni che identificano il sistema di classificazione in uso storicamente o i dati di registrazione/identificazione dei documenti, ecc.) in modo da assicurare l'integrità delle singole unità documentarie e archivistiche e delle relazioni di contesto⁴⁶, ma anche il mantenimento nel lungo periodo in forme stabili delle modalità originarie di reperimento dei documenti e della loro accessibilità, cioè della capacità di comprensione e di elaborazione degli oggetti informatici da parte delle macchine e degli esseri umani. I requisiti funzionali e tecnologici da implementare per la realizzazione dei sistemi documentali includono, inoltre, il rispetto dei principi di conformità alle norme che a livello nazionale stabiliscono i requisiti di validità giuridica dei documenti in forma elettronica e le modalità di autenticazione⁴⁷. Il nodo principale è, quindi, quello della possibilità di contemperare la garanzia dell'integrità e l'esigenza di accessibilità e consentire un "riuso flessibile e illimitato" dei documenti⁴⁸.

Un elemento vincolante è, infine, 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 che determinano, inoltre, un effettivo contenimento dei costi e dei rischi di perdite (in particolare per quanto riguarda la conversione/migrazione delle applicazioni e la duplicazione delle informazioni).

XML sembra offrire 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.

Le potenzialità specifiche di XML in questo ambito riguardano, in particolare, la gestione di documenti e di meta-informazioni indipendenti dal software sia a fini di scambio e ricerca che a fini di conservazione. Come si è visto, uno dei nodi è costituito dalla rappresentazione standardizzata dei documenti⁴⁹ indipendentemente dalle piattaforme di lavoro utilizzate e, quindi in grado di affrontare, il più

⁴⁶ La definizione dei metadati significativi per assicurare l'integrità a lungo termine dei documenti informatici e la loro accessibilità è una questione cruciale per gli archivisti e ha una valenza strettamente teorica, anche se non può prescindere da una attenta valutazione e da un uso adeguato delle tecnologie. Sul tema della definizione dei requisiti per la gestione elettronica del documento (*Model requirements for the management of electronic records*) è al lavoro, finanziato dall'Unione europea nell'ambito del progetto IDA, un gruppo di ricerca che fa capo alla società londinese Cornwell Affiliates e si avvale di esperti internazionali. Anche il citato progetto InterPARES avrà tra i suoi risultati - come si è già ricordato - la identificazione della serie di informazioni necessarie a garantire l'autenticità e l'integrità dei documenti elettronici.

⁴⁷ Nel caso delle amministrazioni pubbliche italiane, l'automazione dei sistemi documentari deve tenere conto di norme europee e di una serie di provvedimenti nazionali molto complessi ancora in via di definizione: dpr 513/97, dpr 428/98, dpcm 8 febbraio 1999, regole tecniche 24/98. Sono in corso di approvazione le regole tecniche applicative del dpr 498/98, mentre non è un caso che siano in fase di prima elaborazione (del tutto insoddisfacente) le disposizioni sulla conservazione dei documenti informatici.

⁴⁸ Enrico Seta, *Digitalizzazione e linguaggi di marcatura*, in "Bollettino AIB", 1999, p. 72.

⁴⁹ Si osservi che nella bozza di regole tecniche predisposte dall'Aipa e dal Dipartimento della funzione pubblica in applicazione del dpr 428/98 sulla gestione informatica dei documenti e in corso di approvazione (cfr <http://www.aipa.it>) l'articolo 16 (leggibilità dei documenti) stabilisce che "ciascuna amministrazione garantisce la leggibilità nel tempo di tutti i documenti trasmessi o ricevuti adottando i formati previsti all'articolo 6, comma 1, lettera b) della delibera Aipa 24/98 ovvero altri formati non proprietari". Il citato articolo 6 fa riferimento in modo specifico ai formati PDF e SGML.

a lungo possibile⁵⁰, i rischi e gli oneri che derivano dalla obsolescenza tecnologica. In questi ultimi anni si è, infatti, assistito a una crescita esponenziale di ambienti applicativi diversi e alla proliferazione dei formati per la creazione di documenti elettronici che a fini conservativi devono essere necessariamente convertiti in prodotti standard in grado di garantire connettività. Tra questi sta ottenendo, non a caso, notevole successo la soluzione fornita dai linguaggi di marcatura, che identificano e mantengono con strumenti indipendenti dall'hardware e dal software metadati strutturati, predefiniti ma allo stesso tempo flessibili, condivisibili e, insieme, suscettibili di un trattamento dettagliato. E' naturalmente indispensabile sviluppare schemi concettuali e grammatiche specifiche per la formazione, gestione e tenuta dei documenti che identifichino le informazioni necessarie al mantenimento e all'uso dei documenti, dai dati di contesto organizzativo a quelli relativi all'ordinamento dei documenti, dalla loro organizzazione in serie e fascicoli al tracciamento degli interventi conservativi, ecc.

L'uso di XML, tuttavia, apre ulteriori e molto significative 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 materiali⁵¹,
- 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 rilevante nei processi di automazione del settore documentario anche in termini 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 Prescod⁵² - 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 innanzi tutto una questione che rimette al centro della progettazione i problemi di struttura logica e concettuale che, naturalmente, richiedono 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. Non è un caso, infatti, che gli ambiti di sviluppo più promettenti si concentrino proprio nella definizione, per ambiti settoriali, di standard internazionali o di procedure normalizzate a livello nazionale. E', ad esempio, il caso dell'Encoded Archival Description promosso dalla Library of Congress⁵³ 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 tra pubbliche amministrazioni⁵⁴.

⁵⁰ E' stato osservato che anche gli standard sono destinati a subire processi di evoluzione e, quindi, di obsolescenza e che per il loro successo e la loro diffusione non è sufficiente l'approvazione da parte degli organismi internazionali, ma servono notevoli investimenti in campo applicativo, tutt'altro che garantiti da un provvedimento ufficiale di riconoscimento.

⁵¹ 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.

⁵² *Ibidem*

⁵³ 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.

⁵⁴ Il provvedimento concluderà 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

*Il progetto di partnership Italia-USA sulla metodologia XML per la conservazione e l'accesso ai documenti elettronici*⁵⁵

Le amministrazioni archivistiche dei Paesi tecnologicamente all'avanguardia esprimono da tempo le loro crescenti preoccupazioni in relazione alla capacità di affrontare adeguatamente il futuro delle memorie digitali. Nel 1998 il responsabile dell'amministrazione archivistica statunitense, John Carlin, sottolineava che la crescita esponenziale dei documenti elettronici prodotti dal governo federale (milioni di file in pochi anni) era ed è incompatibile con le risorse e con gli strumenti disponibili e che il rischio di perdita definitiva in larga parte del patrimonio documentario contemporaneo richiede uno sforzo eccezionale, non solo in termini di investimento per le attrezzature tecnologiche ma anche per la ricerca di soluzioni per la sperimentazione e verifica su larga scala di tecnologie avanzate per la conservazione di documenti elettronici. Da questa preoccupazione era, quindi, nata la decisione del National Archives di Washington di partecipare a un impegnativo programma di ricerca avviato dall'Università della California, il *Distributed Object Computation Testbed (DOCT)*, per valutare soluzioni informatiche avanzate in grado di gestire grandi quantità di documenti digitali⁵⁶. Uno dei punti di forza del progetto NARA-DOCT/Electronic Records Management Project è stato proprio lo sviluppo di strumenti basati sullo standard XML per la migrazione di documenti informatici e dei metadati necessari a garantirne l'accessibilità e a provarne l'integrità.

La prima fase della ricerca che è stata condotta a partire dal 1° ottobre 1998 e che, per quanto riguarda il quadro concettuale di riferimento, è strettamente correlata al progetto InterPARES, ha già dato alcuni primi risultati significativi:

- la definizione di un'architettura scalabile per gestire la migrazione dei supporti,
- l'elaborazione di un modello informativo per trattare la migrazione dei dati di contesto.

La sperimentazione si era, tuttavia, concentrata sul trattamento ai fini della conservazione permanente (nella ricerca si parla di un arco temporale di 400 anni) di un fondo archivistico costituito da oltre un milione di messaggi di posta elettronica conservati presso il National Archives.

La fase successiva che si è aperta alcuni mesi fa grazie a un nuovo finanziamento di 300.000 dollari del National Historical Publications and Records Commission è destinata ad allargare il campo di indagine ad almeno tre grandi classi di documenti elettronici (documenti testuali, documenti composti, documenti GIS) il cui accesso richieda l'uso di strumenti software. Il nodo centrale della ricerca, che corrisponde alla questione di fondo della conservazione delle memorie digitali, è quello di:

- definire un meccanismo per la creazione parzialmente automatica della rappresentazione digitale dei documenti in forme indipendenti dal software e sostitutive di originali che non possono essere conservati a lungo termine per ragioni di obsolescenza,
- predisporre un prototipo di strumento software indipendente dalle piattaforme, sufficientemente robusto, flessibile e scalabile (Archivists' Workbench Software Package), basato sull'utilizzo di XML in quanto standard emergente (e promettente) per la rappresentazione e lo scambio informatico sul web e fondato sui risultati ottenuti nel corso delle precedenti indagini condotte dalla Università della California relative a sistemi di

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. Nella bozza delle regole - che come si è ricordato sono a disposizione sul sito dell'Autorità (www.aipa.it) - 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.

⁵⁵ Il progetto statunitense - NARA-NPACI, "Methodologies for Preservation and Access of Software-dependent Electronic Records" - è stato promosso nel 1998. Una seconda fase di durata triennale del programma di ricerca che ha ottenuto nuovi consistenti finanziamenti dal National Historical Publications and Records Commission è stata approvata nella primavera del 2000 con l'obiettivo specifico di affrontare i problemi di scalabilità delle soluzioni individuate e della loro utilità per ambienti di (<http://www.sdsc.edu/NHPRC>).

⁵⁶ Il progetto è finanziato dall'US Patent and Trademark Office e dalla Defense Advanced Research Projects Agency. Per maggiori informazioni sul progetto cfr <http://www.sdsc.edu/DOCT>.

wrapper-mediator (cioè componenti software che operano come traduttori tra i formati nativi di una fonte informativa e un protocollo comune) anch'essi basati su XML. La scalabilità dei prodotti riguarda la capacità di rispondere anche alle esigenze di depositi archivistici di medie e piccole dimensioni. Un ulteriore sviluppo del progetto riguarda l'integrazione di software esistenti con le funzionalità realizzate con il prototipo.

All'origine di questa scelta c'è la convinzione che i documenti elettronici possano essere considerati come fonti distribuite di informazione *semi-strutturata*, costituite da uno schema definito di componenti informative interne ed esterne al documento e da una serie di elementi passibili di variazione (il supporto, il contesto tecnologico, ecc.).

Il progetto americano si basa su una serie di presupposti e di pre-condizioni:

- la considerazione che la codifica ASCII o Unicode per le informazioni testuali e la codifica bitmap per le immagini siano indipendenti dalle infrastrutture tecnologiche,
- l'assunto per cui la rappresentazione di informazione strutturata mediante linguaggi di marcatura (XML) è indipendente e di facile accesso e consente l'auto-descrizione dei documenti,
- la definizione di una metodologia per la creazione di fonti informative sostitutive degli originali basata sullo sviluppo di "contenitori" (*wrapper*) di prodotti software strutturati in modo che:
 - tutti i metadati che descrivono i contesti documentari abbiano la forma di documenti XML forniti di specifiche DTD,
 - tutte le informazioni testuali siano convertite in documenti XML,
 - tutte le immagini siano convertite in bitmap,
 - tutti i riferimenti a immagini e ad altri documenti all'interno di un documento archivistico siano convertiti in collegamenti permanenti a loro volta rappresentati in un formato XML compatibile.

Un aspetto del progetto che merita una specifica riflessione riguarda la necessità di prevedere modifiche - prodotte anche con procedure automatiche - delle DTD in seguito ad interventi di conversione, migrazione o copiatura dei materiali digitali da parte delle istituzioni archivistiche cui sono affidate.

Alcuni risultati sono già stati raggiunti e riguardano, come si è ricordato, la struttura del modello informativo per la conservazione permanente di materiali archivistici⁵⁷. In particolare, nel progetto si identificano almeno tre nuclei di elementi che devono essere mantenuti nel sistema (simultaneamente alle singole entità documentarie):

- lo schema logico che organizza gli attributi essenziali, cioè
 - i metadati relativi ai documenti singoli (*digital object representation*) che ne definiscono la struttura, il contesto fisico e la provenienza,
 - i metadati che si riferiscono alla organizzazione dell'archivio e includono le diverse informazioni di contesto (*data collection representation*), a loro volta organizzati in sotto-insiemi,
 - i metadati di presentazione (*presentation representation*), che consentono la conservazione di diverse interfacce utente, in particolare dell'interfaccia originaria,
- la descrizione fisica degli attributi all'interno del database del deposito archivistico,
- un dizionario dei dati per le definizioni semantiche degli attributi.

Come emerge anche da questa breve presentazione, i ricercatori sono consapevoli della grande complessità della struttura informativa dell'archivio e delle meta-informazioni che devono essere identificate, mantenute e gestite nel tempo per assolvere il compito della conservazione. Le attività più delicate non riguardano tanto le soluzioni tecnologiche, ma i problemi semantici, ovvero l'individuazione e l'utilizzo delle componenti logiche e la definizione e articolazione dei sotto-sistemi. Perché si ottengano risultati di qualità su questo terreno di ricerca sono necessarie una padronanza dei principi e dei metodi archivistici e una solida esperienza maturata in ambienti, tradizioni e giurisdizioni diverse. E' per questa ragione che gli studiosi statunitensi hanno accolto positivamente la proposta di collaborazione con quelle istituzioni italiane che da anni condividono le medesime preoccupazioni sulla conservazione dei

⁵⁷ Reagan Moore, Chaitan Baru, Arcot Rajasekar, Bertram Ludaescher, Richard Marciano, Michael Wan, Wayne Schroeder e Amarnath Gupta, *Collection-Based Persistent Digital Archives. Part I*, in "D-Lib Magazine", 6 (2000), n. 3, disponibile al seguente indirizzo: <http://www.dlib.org/march00/moore>.

documenti informatici e il medesimo interesse per le potenzialità dei linguaggi di marcatura e che hanno, comunque, avviato un analogo programma di lavoro⁵⁸.

Il progetto italiano, promosso dall'Istituto di studi per la tutela dei beni archivistici e librari dell'Università di Urbino e sostenuto dall'Ufficio centrale per i beni archivistici, dall'Associazione nazionale archivistica italiana e dal Consorzio Roma Ricerche, riguarda in particolare la verifica in ambito europeo dei requisiti funzionali per la conservazione di archivi digitali e la definizione di una metodologia basata sul trattamento dei metadati mediante l'utilizzo di XML e lo sviluppo di DTD in stretta connessione con la ricerca InterPARES⁵⁹ e la complessa indagine NARA-NPACI. L'impegno più significativo - per il quale è prevista la diretta collaborazione con il gruppo di lavoro statunitense - riguarda l'individuazione degli attributi necessari a garantire l'autenticità, l'integrità e l'accessibilità a lungo termine dei documenti elettronici e la loro strutturazione. La collaborazione si basa sull'analisi dei materiali di indagine, sulla comune valutazione del metodo sviluppato e sulla organizzazione congiunta di seminari e workshop.

E' presto per valutare gli esiti di un rapporto appena avviato, anche se sin d'ora si può, comunque, ritenere che l'iniziativa consentirà un confronto molto concreto e operativo su aspetti vitali della ricerca archivistica. Non si può, tuttavia, tacere una considerazione per quanto riguarda le difficoltà in cui si svolge oggi la ricerca in Italia soprattutto in settori che offrono una limitata visibilità: a fronte di consistenti e continui investimenti finanziari da parte delle istituzioni nordamericane, nel nostro Paese il lavoro di indagine è caratterizzato da iniziative quasi individuali sostenute dalle modestissime risorse delle istituzioni culturali (in questo caso l'amministrazione archivistica e l'associazione degli archivisti italiani). Eppure la salvaguardia della memoria documentaria del futuro e i grandi rischi che la minacciano costituiscono un tema vitale per ogni comunità civile che abbia il senso della propria dimensione storica e, almeno, quella della sua continuità. E' vero, purtroppo, che gli orizzonti temporali degli individui e delle amministrazioni si accorciano sempre più e che solo gli specialisti di settore sembrano avere ancora a cuore i problemi - costosi, impegnativi e assai poco remunerativi - della memoria. Per fortuna le tecnologie hanno già più volte dimostrato di saper trovare le risposte anche ai problemi di cui sono esse stesse responsabili. XML è, appunto, uno strumento che apre prospettive incoraggianti per risolvere i rischi di perdita e corruzione dell'informazione digitale.

⁵⁸ In particolare, l'amministrazione archivistica e il Consorzio Roma Ricerche conducono da tempo uno studio e hanno già prodotto alcune realizzazioni sull'uso di SGML/XML per il recupero retrospettivo di strumenti di ricerca archivistici. Si è, ad esempio, recentemente affrontata la digitalizzazione della *Guida generale degli Archivi di Stato italiani* utilizzando il formato XML Cfr <http://www.maas.ccr.it/cgi-win/h3.exe/aguida/findex.it>, con particolare riferimento alle parti intitolate: "La storia della Guida" e "Il progetto informatico".

⁵⁹ Chi scrive, oltre a svolgere le funzioni di direttore dell'Istituto di Urbino, è anche il coordinatore del team italiano che collabora nell'ambito della ricerca InterPARES. Cfr. M. Guercio, *La ricerca InterPARES. Lo stato del progetto*, in "Il mondo degli archivi", 1999, 1, pp. 10-14; Id., *Il futuro per le memorie digitali*, in "Autorità per l'informatica nella pubblica amministrazione, Notiziario", 2000, 1, pp. 50-55; Id., *Qualche informazione sullo stato di avanzamento del progetto Inter-PARES*, in "Il mondo degli archivi", 2000, pp. 47-48.