

Gestione e conservazione di dati e metadati per gli archivi: quali standard?

Pierluigi Feliciati – Università di Macerata

*Metadata are for documents or other information resources as water is to human beings.
Lack of water will lead to dehydration and malfunctioning¹*

*...there are literally hundreds of metadata schemas to choose from
and the number is growing rapidly, as different communities
seek to meet the specific needs of their members²*

Questo intervento non pretende di chiarire in modo definitivo l'attuale quadro in merito agli standard e ai profili applicativi di metadati per la gestione di depositi digitali archivistici, obiettivo che sembra - a chi scrive - perseguibile con pienezza dedicandogli ben altro spazio espositivo e solo da parte di un gruppo di lavoro internazionale. Piuttosto, si vogliono qui raccogliere in modo più possibile ordinato alcune informazioni e riflessioni sul complesso quadro attuale, nell'ottica di un futuro coordinamento – questo, anche solo a livello nazionale - sul tema dei metadati archivistici.

La premessa da cui si potrebbe partire, ovvia per gli addetti ai lavori, costituisce un punto critico nella progettazione di applicazioni documentarie basate sull'*Information and Communication Technology*: non c'è risorsa informativa digitale senza l'accoppiamento di una sequenza di bit con le informazioni sulla rappresentazione che ne è l'origine. Insomma, è fuori discussione l'indivisibilità tra dati/informazioni/documenti e metadati in ambiente digitale³. “Una sequenza (...) di bit è inutilizzabile senza metadati in grado di trasformarla in risorsa” scriveva ormai quasi un decennio fa Giovanni Bergamin⁴, ben evidenziando la differenza tra dati e risorse. Stefano Vitali ha invece definito tutti gli elementi di contesto degli oggetti digitali “soprattutto se devono essere conservati e resi accessibili nel corso del tempo, come entità singole e come reti di significati iscritti in contesti temporali, spaziali, documentari da cui traggono origine e che attraversano nel

1 ERPANET *Training Seminar: Metadata in Digital Preservation - Final report*, Marburg September 3-5, 2003.

2 Chris Taylor, *An Introduction to Metadata*, University of Queensland Library (Australia), 2003, in <http://www.library.uq.edu.au/iad/ctmeta.html>.

3 La definizione di oggetto digitale (data object) in ambito *Open Archive Information System* evidenzia questo assioma: “Oggetto costituito da un insieme di sequenze di bit” che “insieme con le informazioni sulla sua rappresentazione costituisce un oggetto informativo”; vedi *OAIS. Sistema Informativo aperto per l'archiviazione*, a cura di G. Michetti, Roma, ICCU, 2007, *Terminologia*, p. 18.

4 G. Bergamin, *I parenti poveri di Dublin Core: tra modelli e applicazione*, in “Le risorse elettroniche. Definizione, selezione e catalogazione”, atti del convegno internazionale, Roma 26-28 novembre 2001, Editrice Bibliografica, Milano, 2002.

corso del tempo” un loro *doppio* “nel quale specchiarsi, che parli per loro, che ne racconti in qualche modo la storia”⁵.

Inoltre, l'attuale armamentario tecnico per la creazione e gestione di depositi digitali aperti prevede che “sempre più spesso la rappresentazione di un documento” sia “il risultato dell'aggregazione istantanea di componenti digitali distinte all'origine”, anche se a temperare facili ottimismo Giovanni Michetti aggiunge che però “sempre più spesso i sistemi documentari digitali sono purtroppo strutturati in maniera drammaticamente puntiforme”⁶. Insomma, gli archivisti sanno bene che il valore informativo di un documento è praticamente dimezzato se lo si separa dal contesto in cui le attività che ne sono origine lo hanno collocato, e non dimenticano tale principio nel ben più scivoloso e pericoloso ambiente digitale, dove peraltro sono così meno significative le collocazioni fisiche a favore di quelle logiche, rappresentate e garantite, nel tempo, da “dati strumentali di corredo”, secondo un'ulteriore definizione che potremmo dare dei metadati⁷.

Le relazioni tra risorse informative e metadati, dunque, appaiono anche solo ad uno sguardo veloce tanto cruciali quanto particolarmente articolate e dinamiche. Tanto per rammentare alcune tra le proprietà di tali relazioni, basta dire che i metadati possono essere essi stessi statici o dinamici, che possono essere *embedded*, cioè inclusi come parte integrante della risorsa, oppure ad essa collegati sulla base di meccanismi identificativi stabili e sicuri nel tempo, che diversi schemi di metadati con diverse funzioni possono essere associati allo stesso oggetto, che più oggetti possono essere associati tra loro tramite metadati e infine che durante il ciclo di vita di una risorsa è necessario che il *corpus* di metadati sia aggiornato di continuo, per documentare tutti gli eventi ad essa occorsi.

Il tutto, ovviamente, deve anche essere utilizzabile con efficacia, vale a dire identificabile,

5 S. Vitali, *Passato digitale*, Bruno Mondadori, Milano 2004, p. 189

6 OAIS. *Sistema Informativo aperto per l'archiviazione*, cit., *Introduzione*, in part. pp. XI-XII.

7 Le definizioni stesse di oggetti digitali e di metadati presenti nei glossari di riferimento su questi temi confermano l'indissolubilità dei due elementi: *Oggetto digitale*: An entity in which one or more content files and their corresponding metadata are united, physically and/or logically, through the use of a digital wrapper (*California Digital Library Glossary*, <http://www.cdlib.org/inside/diglib/glossary/>). *Documento*: informazione registrata o oggetto che può essere trattato come un'unità (ISO 15489). *Metadati*: dati che descrivono il contesto, il contenuto e la struttura dei documenti e la loro gestione nel tempo (ISO 15489); Informazioni strutturate o semi-strutturate che permettono la creazione, la gestione e l'uso di record nel tempo, e all'interno e attraverso i domini in cui vengono creati (MoReq- Requisiti modello per la gestione di record elettronici). *Recordkeeping Metadata*: Structured or semi-structured information which enables the creation, management, and use of records through time and within and across domains in which they are created. They can be used to identify, authenticate, and contextualize records; and the people, processes and systems that create, manage, and maintain and use them. (Wendy Duff, *Metadata in Digital Preservation: Setting the Scene*, paper for the ERPANET seminar, Marburg 2003, in http://www.erpanet.org/events/2003/marburg/presentations/erpaTraining-Marburg_Duff.pdf)

databile, ricercabile, associabile, restituibile, sia strumentalmente all'uso appropriato delle risorse, da parte della “comunità designata”, sia al fine di garantire funzioni – interne o esterne al deposito - più spiccatamente gestionali.

A rendere ancor più difficilmente governabile questo intreccio tra dati e “informazioni di servizio” interviene la varietà di schemi e modelli di metadati disponibili, sviluppati nell'ambito di contesti e di progetti anche molto diversi, finalizzati a svolgere funzioni simili o identiche, con sovrapposizioni non sempre compatibili. Inoltre, anche se l'adozione pressoché ormai scontata del meta-linguaggio di marcatura XML e del coerente modello concettuale offerto da OAIS (*Open Archival Information System*)⁸ premono per l'inevitabilità dello scambio/condivisione delle risorse informative, restano tuttora ampi spazi di incomunicabilità, tanto che la raffinata attività di *mapping* necessaria a far comunicare tra loro modelli e schemi differenti sembra essere uno dei principali sport praticati dalla comunità dell'*information science*.

Un'altra questione importante da porre, mi pare, è cosa si intenda propriamente per *standard*, se ci si riferisce al variegato panorama di modelli, profili applicativi, formati, linguaggi, protocolli, e in genere degli schemi di riferimento più o meno applicativi delle scienze dell'informazione. Da che momento in poi, e tramite quali meccanismi consuetudinari e/o normativi un modello tecnico opportunamente documentato diviene uno standard? So bene che la domanda è retorica, perché non è certo possibile dare una risposta univoca, però qualche banale considerazione può comunque essere messa in fila. Intanto, pur esistendo degli organismi di certificazione degli standard riconosciuti internazionalmente, e a diversi livelli, come W3C, ISO, NISO, UNI (che peraltro non adottano tutti la stessa politica per la diffusione pubblica della documentazione), non sono certo gli unici a rilasciarne. Si pensi, ad esempio, agli standard prodotti dalla *Library of Congress* statunitense, un'istituzione talmente meritoria e autorevole a livello globale da costituire un punto di riferimento indiscutibile nel panorama delle scienze dell'informazione. Oppure ai formati rilasciati da associazioni tecniche e imprenditoriali, come ad esempio il *Moving Pictures Expert Group*⁹, che ha rilasciato - tra l'altro – schemi di metadati come il

8 Per cui vedi *OAIS. Sistema informativo*, cit.

9 Per la precisione, il MPEG (designazione formale ISO/IEC JTC 1/SC 29/WG 11) è un “un comitato tecnico congiunto delle organizzazioni internazionali ISO e IEC [*International Electrotechnical Commission*] incaricato di definire standard per la rappresentazione in forma digitale di audio, video e altre tipologie di contenuti multimediali in modo da soddisfare un'ampia varietà di applicazioni”, cfr. la voce su Wikipedia in: <http://it.wikipedia.org/wiki/MPEG> e il sito <http://www.chiariglione.org/mpeg/>.

MPEG-21 (ISO/IEC 21000-N)¹⁰, che si occupa della definizione dei requisiti tecnici fondamentali degli oggetti digitali (*digital item*, nell'ottica di questo standard) e della possibilità di interazione da parte dell'utente con i medesimi oggetti. Un passo in più a partire da questo modello è stato compiuto con lo standard strutturale MPEG-21 DIDL, ma se ne dirà più avanti.

Infine, per venire alla realtà italiana, la situazione sembra ancora più fluida, forse a causa di una certa nostra difficoltà ad istituire autorità dedicate al rilascio di schemi, modelli o terminologia (e/o a riconoscerne il ruolo unificante). Si considera, ad esempio, uno standard il profilo applicativo MAG, elaborato dall'Istituto Centrale per il Catalogo Unico del Ministero per i beni e le attività culturali come modello uniforme per i tanti progetti di digitalizzazione sviluppati nell'ambito del programma della Biblioteca Digitale Italiana e di cui si tratterà più avanti. Standard perché schema definito e mantenuto da un comitato ad hoc sostenuto da un istituto centrale, perché applicato da un numero ingente di istituzioni nazionali riconosciute, perché base per lo sviluppo di software specifici da parte di diverse aziende, perché documentato pubblicamente. In modo simile, ma con caratteristiche differenti, ci si rapporta al PICO – Portale Italiano Cultura Online¹¹, uno schema XML basato su *Dublin Core* dedicato a riunire in un modello descrittivo le informazioni di base sui beni e le attività culturali in tutte le loro manifestazioni, fisiche e non: schema documentato, mantenuto, riconosciuto, ma in questo caso applicato attualmente solo per un progetto, per quanto importante e unificante, il portale *Culturaitalia*.

Per quanto riguarda il settore archivistico italiano, la indiscutibile debolezza dell'Istituto Centrale per gli Archivi, rispetto ad esempio alla Direzione per gli Archivi da cui pure dipenderebbe oppure rispetto alle Regioni più attive nel settore, è causa di un certo imbarazzo ogniqualvolta ci si trovi a dover definire schemi terminologici, liste d'autorità, modelli di riferimento, schemi etc. che abbiano valore inter/nazionale. Nell'era del nostro processo di federalismo (confusamente) accelerato, questo fenomeno appare - a chi scrive - solo un altro brutto segno dei tempi.

Per chiudere su questo tema, sembra di poter dire - almeno - che perché si possa parlare

10 “MPEG-21 aims at defining a normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain. This open framework will provide content creators, producers, distributors and service providers with equal opportunities in the MPEG-21 enabled open market”. Cfr. <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>.

11 La documentazione tecnica alla base del complesso progetto Culturaitalia-PICO in <http://www.culturaitalia.it/pico/FootMenu/documentazione/it/index.html>, il *Profilo Applicativo Pico - versione 1.0*, a cura della Scuola Normale Superiore di Pisa, in <http://www.culturaitalia.it/pico/documenti/picoap/picoap1.0.xml>.

di uno standard, per i metadati, è necessario che una comunità sufficientemente ampia e autorevole accolga un modello documentato e lo sperimenti e applichi, evidenziandone eccellenze e criticità, così da correggerlo e mantenerlo aggiornato rispetto a eventuali nuove esigenze.

Venendo più attentamente alle funzioni e ai livelli d'applicazione di metadati per un archivio digitale, la specificità e complessità della questione è stata più volte autorevolmente rimarcata¹² ma può essere utile in questa sede elencare sinteticamente alcuni tra i molti punti critici: prima di tutto, dato che la conservazione a lungo termine degli archivi deve garantire non solo l'accesso, ma anche l'identità e l'integrità dei documenti e delle loro relazioni, è necessario adottare, accanto a modelli e schemi gestionali di livello generale, schemi dettagliati e differenziati per le singole tipologie documentarie e per i contesti applicativi, con la massima attenzione a documentare e conservare, oltre che le risorse documentarie, tutto ciò che documenta le scelte tecnico-organizzative e gli schemi adottati. E questo va coniugato, diacronicamente, anche nel tempo futuro, perché ogni modifica delle procedure che originano le unità archivistiche può modificare i profili documentari e il contesto organizzativo.

Le funzioni dei metadati per la conservazione degli archivi digitali vanno quindi dall'identificazione permanente degli oggetti e delle loro relazioni, con la documentazione di quanto sia necessario per garantirne nel tempo la piena intelligibilità, alla memorizzazione dei meccanismi – tecnici e procedurali - di formazione, tenuta e conservazione del sistema documentario, alla trama dei privilegi d'accesso e al loro mutare, occasionale oppure regolare, in base alle scadenze normative.

Insomma, l'articolazione delle relazioni tra le unità archivistiche digitali e i propri “doppi” (per usare ancora l'efficace metafora di Vitali) si sviluppa in molte direzioni, che possiamo sintetizzare in uno schema tridimensionale:

- *in orizzontale*, la dimensione tempo, peculiare della sedimentazione degli archivi, la cui corretta formazione e tenuta implica controlli, interventi di aggiornamento e aggiunta di informazioni differenziati in base alle fasi del loro ciclo di vita e delle caratteristiche delle attività poste in essere a cura dei soggetti, anch'essi variabili nel tempo e dotati di

¹² Tra tutti, l'efficace M. Guercio – G. Marinelli, *La conservazione delle memorie digitali: un quadro di riferimento per un progetto applicativo*, in «Archivi & Computer», a. XVII, fasc. 1/07, pp. 30-58, in particolare il paragrafo 2.4.2, alle pp. 46-49.

specifiche autorizzazioni;

- in *verticale*, i record d'archivio, nelle loro strutturazioni, gerarchiche e relazionali, di contesto (anch'esse, come sappiamo, soggette a possibili variazioni);
- la terza dimensione, chiamiamola per intenderci *profondità*, vede in gioco (paralleli, sovrapposti, attivati e disattivati a seconda del bisogno) gli schemi/profilo di metadati rispondenti alle varie funzioni per i quali si è scelto di adottarli, il cui numero e la cui complessità sono proporzionali alla complessità delle procedure (descrittive, tecniche, gestionali, di scambio-trasmissione, di conservazione, etc.) e la cui documentazione e aggiornamento è importante ai fini della gestione diacronica dell'*Electronic Records Management System*.

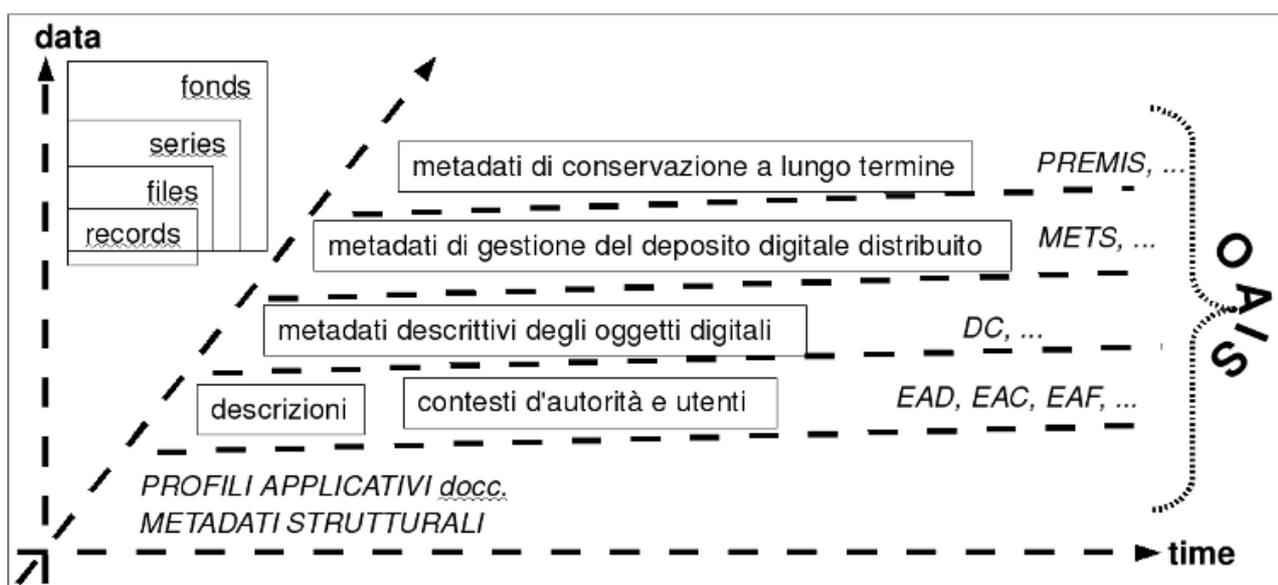


Fig. 1 – le relazioni tra archivio digitale e metadati in tre dimensioni

Per concludere questa rassegna di premesse, credo sia necessario far affiorare alcune ambiguità nell'uso della terminologia tecnica correlata, visto che nei paragrafi seguenti ad essa dovremo rivolgerci per il nostro breve excursus. Quando si parla di schemi di metadati (*metadata schema*) ci si riferisce alla documentazione che descrive un insieme di elementi e le loro interrelazioni, definendo per essi una sintassi formale e strutturale e specifiche regole semantiche (o delle definizioni)¹³. In ambito *Dublin Core*, poi, si è anche affermato che le dichiarazioni degli elementi devono essere rappresentate utilizzando i

13 La definizione dello standard ISO 23081-1:2006 *Metadata for Records* recita: "Schemas are a framework that specifies and describes a standard set of metadata elements and their interrelationships. Schemas provide a formal syntax (or structure) and semantics (or definitions) for the metadata elements". Per lo standard ISO cfr. http://www.iso.org/iso/catalogue_detail.htm?csnumber=40832.

linguaggi XML o RDF¹⁴, così da poter essere processate da agenti software.

A volte può anche essere utile suddividere un singolo schema di metadati in più *file*, specie nel caso di schemi particolarmente complessi o estesi: in questi casi sarà necessario costruire anche uno “schema degli schemi” che garantisca la sussistenza delle relazioni e delle regole che si sono stabilite.

Di altro si tratta quando si usa il termine profilo applicativo (*application profile*)¹⁵ intendendo invece un assemblaggio di elementi selezionati da più schemi differenti, combinati in uno schema nuovo, in una sorta di pacchetto finalizzato a garantire per un determinato contesto applicativo piena modularità ed estensibilità per le funzioni previste, ma senza perdere l'opportunità di interoperare con altre applicazioni che adottano gli schemi di metadati origine. Il processo di definizione di un *application profile* non esclude l'aggiunta *ex novo* di specifici elementi (racchiusi in un nuovo schema dichiarato) utili ad una specifica comunità o ad un'organizzazione, non necessariamente significativi in contesti differenti da quello per cui sono concepiti¹⁶.

In ambito archivistico digitale, si parla di profili applicativi riferendosi perlopiù all'insieme degli elementi descrittivi necessari a restituire tutte le informazioni sui record in formazione e sui loro contesti¹⁷, da interfacciare con altre informazioni utili alla gestione nel tempo dell'*Electronic Records Management System*. La corretta sottolineatura che si fa di norma della necessità di “cucire su misura” i profili applicativi documentari rispetto alla complessa stratigrafia dei contesti organizzativi e istituzionali, delle tipologie dei soggetti produttori e

14 DCMI *tools glossary*, in <http://dublincore.org/groups/tools/glossary.shtml>.

15 “An assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema. Application profiles provide the means to express principles of modularity and extensibility. The purpose of an application profile is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schemas. Part of such an adaptation may include the elaboration of local metadata elements that have importance in a given community or organization, but which are not expected to be important in a wider context”. Anche qui, la fonte è il DCMI *tools glossary*, cit.

16 Si veda anche l'utilissimo R. Heery and M. Patel. *Application profiles: mixing and matching metadata schemas*, <http://http://www.ariadne.ac.uk/issue25/app-profiles/>. Si rinvia inoltre al fronte di attività Dublin Core denominato *The Singapore Framework for Dublin Core Application Profiles*: “for designing metadata applications for maximum interoperability and for documenting such applications for maximum reusability. The framework defines a set of descriptive components that are necessary or useful for documenting an Application Profile and describes how how these documentary standards relate to standard domain models and Semantic Web foundation standards. The framework forms a basis for reviewing Application Profiles for documentary completeness and for conformance with Web-architectural principles”, in <http://dublincore.org/documents/2008/01/14/singapore-framework/>.

17 Così ad esempio Federico Valacchi, *La memoria integrata nell'era digitale. Continuità archivistica e innovazione tecnologica*, Titivillus, Pisa 2006, a p. 116: “il profilo applicativo ha dunque il ruolo di individuare e catturare tutti gli elementi descrittivi dalla cui combinazione dovrà scaturire la ricostruzione delle diverse componenti che costituiscono l'archivio (...) Insieme ai documenti si dovranno poi descrivere tutte le altre componenti del sistema archivio (...), anch'esse gestite come entità informative autonome da ricondurre poi all'interno del sistema di relazioni nel suo complesso”.

infine delle tipologie documentarie funzionali alle attività di questi ultimi evoca scenari parzialmente differenti rispetto a quanto si è detto rispetto ai profili applicativi standard di metadati e al loro mantenimento. Prima di tutto, perché i profili archivistici devono attingere ai modelli e agli standard tecnici e informativi - che garantiscano interoperabilità, corretta gestione e conservazione dei sistemi distribuiti – oltre che essere totalmente *compliant* rispetto alla normativa in vigore e hanno quindi come criticità principale la necessità di un controllo continuo che ne assicuri la rispondenza al contesto applicativo. Questo controllo va effettuato in base a criteri, da parte di professionalità e con tempistica che sono del tutto differenti rispetto ai profili applicativi standard, la cui dinamica segue i ritmi della tecnologia e dei grandi progetti internazionali e non quelli della vita dei soggetti produttori e dei loro archivi.

Il profilo applicativo adottato da un archivio digitale, insomma, non può essere che un “grappolo” di profili e di schemi combinati, attivi in fasi diverse del ciclo di vita dell’archivio, di cui alcuni – quelli più vicini allo strato descrittivo - saranno provenienti direttamente dall’esperienza archivistica, mentre altri – più tecnici - saranno dedicati a garantire stabilità, autenticità e integrità degli oggetti digitali/documenti e delle loro relazioni, la gestione sicura e a lungo termine dei depositi digitali, l’identificazione degli utenti e la gestione degli accessi, mentre altri ancora saranno finalizzati ad amministrare la trasmissione e lo scambio dei dati tra più sedi fisiche.

Volendo ragionare di schemi di metadati¹⁸, allora, è necessario sottolineare che la maggior parte degli standard che vanno attualmente per la maggiore sono nati in seno a due comunità, quella dei bibliotecari – da sempre interessati agli schemi di mediazione informativa diversi dal linguaggio naturale e decisamente avanzati nell’edificazione di sistemi informativi cooperativi e interoperabili – e quella degli informatici, che considerato quanto si è già detto sopra circa la natura doppia degli oggetti digitali hanno dovuto fissare regole condivise a livello internazionale per identificare, descrivere, gestire e scambiare le sudate e preziose sequenze di bit. Ovviamente, molti dei progetti da cui sono emersi gli schemi che hanno dimostrato maggiore efficacia e durata nel tempo sono stati elaborati da gruppi di lavoro *meticci*, mettendo a frutto esperienze e competenze provenienti da diverse comunità scientifiche.

18 Una buona presentazione generale delle problematiche relative ai metadati, per quanto datata, è quella di Chris Taylor, *An Introduction to Metadata*, cit.

Dal primo “mondo” ci arrivano gli standard dedicati a identificare, descrivere e gestire i “contenitori” digitali, che non a caso vengono battezzati - pur con alcune varianti – adottando termini ispirati al mondo delle biblioteche e talvolta adottati anche in altri ambiti: *digital library*, *digital repository*, collezione, etc.¹⁹. Oppure, si parla spesso con scioltezza in ambito ICT e bibliotecario di *archivi digitali*, riferendosi a qualsiasi aggregazione di oggetti digitali a scopo di conservazione, dunque con un significato ben diverso da quello che qui si è dato e contribuendo ad alimentare un’ambiguità che sembra sempre più difficile sciogliere²⁰.

Sull’identificazione sicura e stabile degli oggetti digitali in ambiente di rete molto ci sarebbe da dire, premettendo comunque che per via delle caratteristiche dell’attuale *Internet governance* sono oggi in vigore contemporaneamente troppi standard “globali”, basati o meno sull’accoppiata storica DNS-URL, coll’effetto di vanificare l’efficacia universale delle funzioni identificative, indispensabili perché la Grande Rete, nella sua versione più avanzata di sistema dei sistemi informativi, sia davvero ambiente di interoperabilità negli ambiti scientifici, culturali e anche dell’amministrazione²¹. Gli standard di identificazione non sono comunque standard di metadati in senso proprio, anche se ne costituiscono una premessa fondamentale ma tutt’altro che pacifica, sussistendo alcune questioni esiziali sulla effettiva granularità delle risorse identificate, sovente aggregatori di più risorse che vengono rappresentate come fossero una sola²².

Tornando ai metadati, va anche sottolineato che, a seconda dell’uso che si fa di dati e metadati, la natura dei due concetti (e i modelli standard corrispondenti) tende talvolta a confondersi: “risulta solitamente chiaro che i dati di indicizzazione essenziali relativi ad un record (titolo, data, ecc.) fanno parte dei metadati del record in questione. Tuttavia, la *audit*

19 Interessante quanto precisa in merito Mariella Guercio, *Gli archivi come depositi di memorie digitali*, in «DigItalia», numero 2 – 2008, pp. 37-54, in part. p. 44; si veda però in proposito anche Anna Maria Tammaro, *Che cos’è una biblioteca digitale?*, in «DigItalia», numero 0 – 2005, pp. 14-33.

20 Un’evocazione della tendenza a confondere la nozione di archivi con quella di “raccolta di dati e informazioni” in Maria Pia Rinaldi Mariani, *Convergenze e mediazioni*, in «Archivi & Computer», 3/2002, p. 18.

21 Un’efficace sintesi sul tema in M. Sebastiani, *Identificatori persistenti per gli oggetti digitali*, in «DigItalia», numero 0 – 2005, pp. 62-84.

22 A questo proposito, non strettamente in tema rispetto a questo intervento ma che ritengo di grande importanza per la gestione delle risorse digitali in rete, si rinvia ad esempio ai concetti alla base dell’OAI-ORE: *Open Archives Initiative - Object Reuse and Exchange*. Nell’*ORE User Guide - Primer* (17 October 2008), in <http://www.openarchives.org/ore/1.0/primer>. Nel par. 2., *Motivating Example*, si chiarisce infatti bene come gli URI si riferiscano di norma alla “human start page” di un documento sul Web, che in realtà risulta ben più articolato in quanto a struttura interna, relazioni con altri documenti, formati, versioni, responsabilità e dinamica nel tempo. Se l’utente-umano può riuscire abbastanza bene a districarsi, non altrettanto può un agente automatico, a meno che non gli si forniscano le euristiche specifiche del contesto da “comprendere”, funzione per cui OAI-ORE è stato concepito.

trail o il programma di conservazione relativi ad un record archivistico, possono essere legittimamente considerati dati o metadati, in funzione del contesto. Possono essere definiti, ad esempio, vari tipi di record per l'indicizzazione, la conservazione, la riproduzione, ecc.”²³. Il forte ancoraggio dei metadati al contesto in cui vengono adottati costituisce dunque irrinunciabile punto di partenza, da cui deriva tra l'altro la necessità di documentare diacronicamente tanto i dati che i metadati, possibilmente "organizzando" questi ultimi in profili applicativi per semplificarne l'uso e la gestione/conservazione nel tempo. Il modello OAIS può essere usato – tra l'altro - per individuare il sistema di metadati adottato e costruire così un sistema di elementi informativi funzionale alla conservazione a lungo termine.

Altra questione spinosa è comprendere la tipologia funzionale dei metadati, sempre meno facile da distinguere. E' ormai invalso l'uso di una classificazione di massima che, vista la frequente sovrapposizione tra le funzioni degli schemi adottati per molti dei profili applicativi "standard", non può certo essere considerata come stabile e definitiva, ma resta utile per l'orientamento tra modelli²⁴. Tra le tassonomie, quella adottata più di frequente e che qui si proverà a seguire a scopi, appunto, espositivi, è quella di Wendler²⁵, che distingue i metadati in tre categorie funzionali:

- “*Descrittivi*: per l'identificazione ed il recupero degli oggetti digitali; sono costituiti da descrizioni normalizzate dei documenti fonte (o dei documenti digitali nativi), risiedono generalmente nelle basi dati dei sistemi di *Information Retrieval* all'esterno degli archivi degli oggetti digitali, e sono collegati a questi ultimi tramite appositi link;
- *amministrativi e gestionali*: per le svariate operazioni di gestione degli oggetti digitali all'interno dell'archivio; [comprendono di norma anche i *metadati tecnici*, relativi agli oggetti digitali nella loro natura più strettamente tecnica (formati, mezzi e procedure di creazione, elaborazione, post-elaborazioni, etc.)];

23 Tale osservazione è tratta da *MoReq - Requisiti modello per la gestione di record elettronici . Specifiche MoReq*, elaborate da Cornwell Affiliates plc per il programma IDA della Commissione europea, Bruxelles-Lussemburgo, 2001, cap. 2.1 *Terminologia chiave*, p.7, in <http://www.europa.eu.int/ispo/ida>. Per ulteriori materiali sui metadati in ambito archivistico si veda anche la documentazione dell'*Archiving Metadata Forum*, in <http://www.archiefschool.nl/amf>.

24 Alla domanda “Which elements, sub-elements and schemes should I use?” Chris Taylor, *An Introduction to Metadata*, cit. risponde saggiamente: “There is no simple answer to this question. At a fundamental level, it becomes a compromise, based on: the specific needs of the local community to maximise information retrieval and management; the need to guard against making the creation of metadata and its maintenance more trouble than it is worth and therefore defeating its purpose; sustainability of the metadata schema in terms of keeping the records up to date”.

25 R. Wendler, *LDI Update: Metadata in the Library*, in: “Library Notes”, n. 1286 (1999), pp. 4-5.

- *strutturali*: per descrivere la struttura interna dei documenti (es. introduzione, capitoli, indice di un libro) e gestire le relazioni fra le varie parti componenti degli oggetti digitali”²⁶.

Di questi, i primi sono quelli più simili agli schemi su cui si basano le classiche attività di descrizione, inventariazione o catalogazione riguardanti oggetti analogici, i secondi somigliano agli elementi gestionali che pure sono presenti in alcune tipologie di descrizioni dell'analogico (riferimenti topografici, numeri di catalogo, riferimenti alla responsabilità della descrizione, etc.), infine i terzi sono separati dai primi solo per le specifiche proprietà dell'ambiente digitale, mentre nel mondo del tangibile la struttura o è parte integrante della descrizione (ad esempio per gli archivi o per gli spogli delle riviste), o non è affatto (ad esempio nell'articolazione in capitoli o pagine di un oggetto bibliografico, che non è oggetto di catalogazione).

In ambito archivistico, si preferisce classificare i metadati come:

- *metadati identificativi* (per il soggetto produttore, le unità documentarie e archivistiche e per i formati elettronici abilitati);
- *metadati sul contesto storico-istituzionale* (per il soggetto produttore);
- *metadati sul contesto organizzativo* (per il soggetto produttore);
- *metadati sul contesto archivistico* (per il soggetto produttore);
- *metadati sul contesto tecnologico* (per il soggetto produttore);
- *metadati sugli oggetti informativi digitali*²⁷.

Se si confronta questo elenco con la tassonomia di Wendler, i metadati identificativi dei formati elettronici e quelli sugli oggetti informativi digitali rientrerebbero in quelli descrittivi e in quelli gestionali/tecnici, adottando ovviamente diversi profili, mentre tutti gli altri sono relativi a quella che chiamiamo usualmente descrizione archivistica, per cui è necessario adottare uno o più profili articolati, funzionali alla gestione e conservazione di un archivio digitale in senso proprio ma che siano più possibile compatibili con EAD²⁸ ed EAC-CPF²⁹.

26 Si riporta qui la tassonomia di Wendler come presentata nel *Manuale MAG: ICCU-Comitato MAG, MAG Metadati Amministrativi e Gestionali. Manuale Utente*, a cura di Elena Pierazzo, ICCU, Roma 2006, pp. 7-8.

27 Cfr. S. Pigliapoco, *Le fasi del processo di conservazione digitale*, in «Archivi & Computer», anno XVII, fasc. 1/07, pp.97-105, in part. alle pp. 100-103.

28 *Encoded Archival Description*, version 2002, “a nonproprietary encoding standard for machine-readable finding aids such as inventories, registers, indexes, and other documents created by archives, libraries, museums, and manuscript repositories to support the use of their holdings.” su cui cfr. <http://www.loc.gov/ead/>.

29 *L'Encoded Archival Context – Corporate bodies, Persons and Families*, solo recentemente rilasciato in una versione draft, "is a standard for encoding contextual information about persons, corporate bodies, and families related to

In aggiunta alle categorie di metadati finora elencate si parla anche di metadati per la gestione dei diritti³⁰, di metadati per la security, di metadati per le informazioni personali, di metadati per la conservazione a lungo termine ed di altri ancora, tutte funzioni queste cui sono stati dedicati diversi schemi specifici, ai più interessanti dei quali per il contesto archivistico digitale avremo modo di accennare più avanti.

Accennando agli standard di metadati descrittivi degli oggetti digitali, non si può non partire dal progetto del Dublin Core (il cui nome completo è *Dublin Core Metadata Initiative*): nel marzo 1995 si tenne una conferenza nella città americana di Dublin (Ohio), durante la quale i partecipanti – bibliotecari, archivisti, editori, ricercatori e sviluppatori di software e rappresentanti dell'*Internet Engineering Task Force*³¹ – convennero sulla necessità di definire strumenti standard per consentire l'accesso alle risorse digitali in rete, a partire da un insieme base di elementi descrittivi, forniti dall'autore o dall'editore dell'oggetto digitale, ed inclusi nell'oggetto o ad esso referenziati stabilmente³².

Lo standard DC-Dublin Core, per la sua semplicità, le sue potenzialità cross-domain e il suo successo, è stato recentemente riconosciuto come standard ISO (15836:2009) e NISO (Z39.85-2001), sottolineando che esso “defines the elements typically used in the context of an application profile which constrains or specifies their use in accordance with local or community-based requirements and policies”³³ (mie le sottolineature), cioè che usato da solo, vale a dire non inquadrato in uno specifico profilo applicativo e adottato da una comunità in base a precise politiche, serve a ben poco.

Citando poi gli altri standard descrittivi del digitale, i *meta tag* previsti per essere inclusi nella sezione HEAD dei file HTML³⁴, sembrano ormai insufficienti rispetto allo schema di 15 elementi di Dublin Core, soprattutto perché riferiti necessariamente alla

archival materials using Extensible Markup Language (XML). The standard is maintained by the Society of American Archivists in partnership with the Berlin State Library and the German Bundesarchiv (Federal Archives)". Cfr. <http://eac.staatsbibliothek-berlin.de/>. In proposito cfr. anche il recente e stimolante contributo di Giovanni Michetti, *EAC. Elementi per un Approccio Critico*, in «Archivi & Computer», a. XVIII, fasc. 1/08, pp. 40-55.

30 Per orientarsi sul complesso framework dei diritti d'uso e di accesso degli oggetti digitali, rinvio alle ontologie per la gestione dei diritti prodotte nell'ambito di CASPAR, in <http://www.casparpreserves.eu/publications/ontologies/RightsOntology> e a ODRL - *Open Digital Rights Language*, in <http://odrl.net/1.1/ODRL-11.pdf>, a cura di Renato Iannella.

31 “The mission of the IETF is make the Internet work better by producing high quality, relevant technical documents that influence the way people design, use, and manage the Internet”, cfr. <http://www.ietf.org/> .

32 Su Dublin Core cfr. <http://www.dublincore.org/documents/dces/> di cui l'ICCU ha curato la traduzione italiana delle specifiche descrittive, in <http://www.iccu.sbn.it/dublinco.html>.

33 Cfr. la presentazione dello standard in http://www.iso.org/iso/catalogue_detail.htm?csnumber=52142

34 Su cui cfr. ad esempio il *Dictionary of HTML META Tags* di Andrew Vancouver in: <http://vancouver-webpages.com/META/> .

rappresentazione web di una o più risorse e non al singolo oggetto digitale. Non mi pare utile invece una comparazione di DC rispetto a schemi altrettanto usati ma concepiti per assolvere a funzioni più specifiche, come MARC - *MAchine-Readable Cataloging* (*Library of Congress* e *Library and Archives Canada*), ad esempio, oppure ISO19115 – *Metadati* (parte degli standard prodotti da ISO/TC211³⁵) che definisce gli schemi necessari per descrivere le informazioni geografiche ed i servizi, e fornisce informazioni relative all'identificazione, l'estensione, la qualità, gli schemi temporali e spaziali, i riferimenti spaziali e la distribuzione di dati geografici numerici. Ancora, lo standard NISO *Metadata for Images* (comunemente noto come NISO MIX)³⁶, sempre sviluppato della Library of Congress, è uno schema di codifica per la descrizione e gestione delle *still images* digitali, a lungo tenuto in status di draft e solo recentemente rilasciato in una versione stabile.

Va anche ricordato che OAI-PMH³⁷, l'efficiente protocollo per la raccolta/allineamento di metadati in sistemi distribuiti basato sul modello OAIS, su cui sono basati i più importanti software attuali per repository istituzionali (come *DSpace*, *ePrints*, *Fedora* e *Opus*), prevede l'adozione di Dublin Core – o comunque di uno schema in cui l'unità di informazione sia un singolo oggetto – per lo schema descrittivo dei record³⁸. I tentativi di adottare schemi descrittivi, pur espressi in XML, più complessi di Dublin Core (che seppur *extended* si riferisce sempre a singoli oggetti e non a strutture), come ad esempio, EAD o TEI-XML³⁹ hanno dimostrato alcune difficoltà applicative ancora non risolte pienamente⁴⁰.

Con caratteristiche simili a Dublin Core, invece, ma nato in seno alla Library of Congress e più legato alle attività di catalogazione bibliografica, va citato lo standard MODS –

35 Cfr. <http://www.isotc211.org/> e in particolare il recente http://www.isotc211.org/Outreach/ISO_TC%20211_Standards_Guide.pdf.

36 Cfr. <http://www.loc.gov/standards/mix//> e <http://www.niso.org/pdfs/DataDict.pdf>.

37 *The Open Archives Initiative Protocol for Metadata Harvesting*, version 2.0, su cui cfr. <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

38 “At a minimum, repositories **must** be able to return records with metadata expressed in the Dublin Core format, without any qualification. Optionally, a repository **may** also disseminate other formats of metadata” recita il paragrafo 2.5 (Record) del *Protocol for Metadata Harvesting*, version 2.0, cit.

39 La Text Encoding Initiative (TEI) è uno standard internazionale nato per permettere alle biblioteche, ai musei, alle case editrici ed a tutti gli studiosi di rappresentare fedelmente qualsiasi testo in formato digitale; lo standard è gestito da un consorzio di istituzioni internazionali di ambito linguistico e letterario (*l'Association for Computers and the Humanities*, *l'Association for Computational Linguistics* e *l'Association for Literary and Linguistic Computing*). Cfr. <http://www.tei-c.org/> e *Il manuale TEI Lite. Introduzione alla codifica elettronica dei testi letterari*, a cura di Fabio Ciotti, Sylvestre Bonnard, Milano 2005, e in http://www.tei-c.org/Guidelines/Customization/Lite/teiu5_it.xml.

40 In proposito si rinvia all'ottimo Muriel Foulonneau - Francis André, *Investigative Study of Standards for Digital Repositories and Related Services*, Amsterdam University Press, Amsterdam 2008, frutto dei risultati del progetto DRIVER – *Digital Repository Infrastructure Vision for European Research*, per cui vedi <http://www.driver-repository.eu/>. Le osservazioni sull'adozione per OAI-PMH di schemi di metadati diversi da DC-Simple alle pp. 27-30.

*Metadata Object Description Set*⁴¹, pensato specificamente per gli oggetti digitali in rete semplificando lo standard bibliografico MARC, di cui eredita alcuni elementi e che vuol consentire la convergenza tra descrizioni XML MARC e non. Lo schema, registrato negli USA come standard NISO è stato pensato anche per fungere da schema descrittivo da includere in METS, di cui si dirà più avanti, e si affianca al *Metadata Authority Description Schema* (MADS)⁴² dedicato alla descrizione di record d'autorità.

Infine, non sono da dimenticare le diverse esperienze di definizione di schemi per la descrizione di collezioni digitali, cioè di aggregazioni sufficientemente stabili di oggetti digitali legati tra loro da vincoli stabili, più o meno volontari, intorno alle quali vengono di norma costruiti i servizi, anche di molti tipi diversi. Senza entrare troppo nel merito di questo fronte, che pare a chi scrive particolarmente interessante per le sue potenzialità di contestualizzazione di ogni oggetto nell'ambito di un quadro che include necessariamente anche i soggetti, i contesti applicativi e le tipologie dei servizi, basta elencare il modello descrittivo dell'anglosassone Research Support Library Programme (RSLP)⁴³, il NISO Metasearch *Collection Description Specification*⁴⁴, il *Dublin Core Collection Description Application Profile*⁴⁵ e il data model relazionale del progetto MICHAEL (*Multilingual Inventory of Cultural Heritage in Europe*)⁴⁶.

Va infine chiarita, a scanso d'equivoci, la profonda differenza tra gli schemi XML di metadati descrittivi fin qui citati e gli schemi XML descrittivi archivistici, come EAD ed EAC-CPF: questi ultimi, ben più articolati strutturalmente e approfonditi semanticamente, non sono concepiti propriamente per descrivere singoli oggetti digitali, pur archivistici che siano, ma a restituire descrizioni archivistiche – sulla base sostanzialmente degli standard descrittivi archivistici ISAD(G), ISAAR(CPF) e ISAF⁴⁷ - rispettando struttura e sistema di relazioni degli archivi e dei soggetti coinvolti. Inoltre, seppure si possano utilizzare gli schemi EAD e EAC-CPF come base per strutturare un archivio digitale in formazione, neanche in questo caso si potrà parlare strettamente di metadati descrittivi, a meno che non si ci si riferisca agli elementi descrittivi - funzionali al collegamento dinamico tra

41 <http://www.loc.gov/standards/mods/>; il MODS *Implementation registry* in <http://www.loc.gov/mods/registry.php> .

42 <http://www.loc.gov/standards/mads/> .

43 <http://www.ukoln.ac.uk/metadata/rslp/>.

44 <http://www.niso.org/workrooms/mi/Z39-91-DSFTU.pdf> .

45 <http://dublincore.org/groups/collections/collection-application-profile/>.

46 Il modello dati, nella traduzione italiana, in: http://www.michael-culture.eu/documents/MICHAELDataModel_ITA.pdf . Per i risultati della campagna di censimento a livello europeo delle collezioni digitali culturali, cfr. il portale europeo <http://www.michael-culture.org/>.

47 Cfr. la pagina degli standard del sito web dell'*International Council of Archives*, in: <http://www.ica.org/en/standards>.

documenti e oggetti digitali - estratti automaticamente dai file XML contengono la descrizione dei record e dei loro contesti.

Passando alla famiglia degli schemi di metadati amministrativo/gestionali, o standard strutturali come vengono talvolta definiti, visto che comprendono funzioni strutturali e di impacchettamento nell'ottica OAIS, di sicuro il posto d'onore spetta allo standard statunitense METS - *Metadata Encoding and Transmission Standard*, già citato⁴⁸. Sviluppato dalla Library of Congress, mantenuto dal suo *Network Development and MARC Standards Office*, lo schema – che nel 2004 ha ricevuto la NISO Registration, rinnovata poi nel 2006 - è finalizzato a raccogliere metadati descrittivi, amministrativi e strutturali “regarding objects within a digital library”: insomma, sia per la gestione degli oggetti contenuti in un deposito digitale, che per lo scambio di oggetti tra i depositi (o tra i depositi ed i loro utenti). Un documento METS potrebbe essere quindi usato sia come *Submission Information Package* (SIP), sia come *Archival Information Package* (AIP), che come *Dissemination Information Package* (DIP) sempre nel quadro del modello di riferimento OAIS. Il punto di forza di METS sta senza dubbio nella sua flessibilità, essendo (quasi) del tutto indipendente dagli schemi di metadati che “impacchetta”, specie per quanto riguarda quelli descrittivi: infatti, ogni elemento `<dmdSec>` (*Descriptive Metadata Section*) potrebbe contenere sia un puntatore a metadati esterni, sia includere i metadati internamente (in due forme: codificati in XML, con il codice identificativo come appartenente ad un *namespace*⁴⁹ diverso da quello del METS; in qualsiasi arbitraria forma binaria o di testo), oppure ancora contenerli entrambe. Tale flessibilità nell'adozione degli schemi descrittivi e tecnici, indubbiamente un vantaggio per la applicabilità in contesti anche molto diversi tra loro, può tuttavia costituire un elemento negativo dal punto di vista dell'interoperabilità dei sistemi, limitando le opportunità di scambio tra depositi che adottano schemi descrittivi o tecnici differenti.

Gli elementi `<amdSec>` contengono i metadati amministrativi relativi sia ai file che

48 Cfr. <http://www.loc.gov/standards/mets/>. Il documento introduttivo sullo schema “METS: An Overview & Tutorial”, del 2004 è stato tradotto in italiano a cura di Angela Di Iorio nel corso del progetto per la biblioteca digitale (S.I.M.B.A.D.), realizzato per la Biblioteca Provinciale "P. Albino" di Campobasso (Italia) ed è disponibile in: <http://www.loc.gov/standards/mets/METSita.html> .

49 Per *namespace*, in XML, si intende lo strumento per poter usare dentro al proprio documento XML alcuni elementi che hanno una particolare provenienza, vale a dire che sono definiti in un qualche schema di codifica. Un namespace è costituito da un prefisso distintivo, ed è necessario dichiararlo all'interno di un qualsiasi elemento di un documento XML, grazie a un attributo standard speciale, previsto dalle specifiche XML, `xmlns:prefisso` (ad esempio `xmlns:dc`) il cui valore è dato dalla URL dove si trova la documentazione dello schema cui il namespace fa riferimento (quindi <http://purl.org/dc/elements/1.1/>). Cfr. <http://www.w3.org/TR/REC-xml-names/> e *Manuale MAG*, cit. par. 2.3.

costituiscono l'oggetto della biblioteca digitale che quelli relativi ai file usati per creare l'oggetto partendo dal materiale originale di provenienza. METS quattro tipologie principali di metadati: metadati tecnici (informazioni riguardanti la creazione, il formato e le caratteristiche di utilizzo), metadati sulla proprietà intellettuale (copyright e informazioni sulle licenze d'uso), metadati sull'origine analogica degli oggetti digitali (descrittivi ed amministrativi riguardanti l'origine analogica di derivazione dell'oggetto della biblioteca digitale) e infine metadati sulla provenienza digitale (sulle relazioni tra i file sorgente e di destinazione, oppure tra file master e di derivazione e sui file impiegati nella migrazione/trasformazione tra la digitalizzazione originale di un oggetto e la sua "incarnazione" come oggetto della biblioteca digitale). La sezione della mappa strutturale di un documento METS (elemento <structMap>) definisce la struttura gerarchica degli oggetti della biblioteca digitale da presentare all'utente, in modo da permettergli di consultarli. Infine, la sezione comportamento può essere usata per associare comportamenti eseguibili al contenuto dell'oggetto METS e contiene uno o più elementi <behavior>, ognuno dei quali ha un definizione di interfaccia che rappresenta una definizione astratta dell'insieme di comportamenti rappresentati in una particolare sezione. Al momento attuale sono pochi i modelli di metadati *multipurpose* che possano confrontarsi con METS: uno è senza dubbio MPEG-21 DIDL (Digital Item Declaration Language), cui si è accennato più sopra, contenuto nella parte 2 dello standard ISO/IEC 21000-2:2003: uno schema di codifica finalizzato a dichiarare la struttura e le caratteristiche dei *digital item*⁵⁰, le cui maggiori caratteristiche sono la flessibilità e l'interoperabilità.

In Italia, dal 2000 è attivo presso l'ICCU il progetto MAG – Metadati Amministrativi Gestionali⁵¹, che ha prodotto un *application profile* con l'obiettivo di fornire le specifiche formali per la fase di raccolta, di trasferimento e disseminazione dei metadati e dei dati digitali nei rispettivi archivi. Lo Schema MAG, realizzato e mantenuto dal Comitato *ad hoc*⁵², ha visto il suo ambito di più immediata applicazione nei progetti che fanno

50 E' opportuno chiarire che per MPEG-21 i Digital Item "sono oggetti digitali strutturati che includono una rappresentazione standard, un'identificativo e dei metadati. Più concretamente, un Digital Item è costituito di una combinazione di risorse (come uno stream video, tracce audio, immagini statiche, ecc.), metadati (come descrittori, identificativi, ecc.) e strutture (che descrivono le relazioni che intercorrono fra le risorse)". Cfr. *Manuale MAG*, cit., par. 1.4.4.

51 Sul profilo MAG cfr. Cristina Magliano, *Lo standard nazionale dei metadati gestionali amministrativi*, in «DigItalia», numero 0 - 2005, pp. 34-46 e le pagine del sito web dell'Istituto Centrale per il Catalogo Unico raggiungibili a partire da quella del Comitato MAG in: <http://www.iccu.sbn.it/genera.jsp?id=99>.

52 Il Comitato MAG è costituito da: Francesco Baldi (Discoteca di Stato); Giovanni Bergamin (Biblioteca Nazionale Centrale di Firenze); Gianfranco Crupi (Università degli Studi La Sapienza di Roma); Gloria Cirocchi, Simona Gatta

riferimento alla Biblioteca Digitale Italiana (BDI), vale a dire l'insieme delle attività di documentazione digitale che hanno inteso condividere determinati standard e hanno potuto essere per questo certificate.

Nel dettaglio, lo schema MAG è composto di quattro file (metadigit.xsd, metatype.xsd, audio.xsd. e video.xsd), collegati gli uni agli altri mediante a un meccanismo di inclusione, e di quattro diversi *namespace*: *mag*, il target namespace (ovvero il vocabolario dei tipi e delle dichiarazioni di elementi); *niso*, lo schema che traduce le linee guida del Data Dictionary NISO (realizzato dal Comitato MAG in attesa del rilascio di NISO MIX); *dc*, lo schema Dublin Core, distribuito direttamente dal Dublin Core Metadata Initiative; *xlink*, lo schema W3C comprendente una serie di attributi che consentono di creare collegamenti fra risorse diverse. Il file metadigit.xsd è il file principale dello schema MAG, che contiene la dichiarazione dell'elemento *root* dello schema, mentre i tipi MAG, semplici o complessi, sono tutti definiti nei file ancillari metatype.xsd, che definisce i tipi di uso generalizzato e relativi alle sezioni IMG, DOC, OCR e DIS, audio.xsd, che definisce i tipi usati dai file audio documentati nella sezione AUDIO e video.xsd, che definisce i tipi usati dai file video documentati nella sezione VIDEO.

Il file MAG è poi diviso in sezioni e per essere valido deve necessariamente contenere le sezioni marcate dagli elementi <gen> e <bib>, che contengono fondamentali informazioni circa l'istituzione che opera la digitalizzazione, il progetto di digitalizzazione, lo stato dell'oggetto digitale e il codice identificativo dell'oggetto stesso. Invece, sono opzionali e ripetibili (eccetto <dis>) le sezioni marcate dagli elementi <stru> (contiene metadati strutturali dell'oggetto digitale), (metadati specifici per le immagini fisse), <audio> (metadati specifici per file audio), <video> (metadati specifici per file video), <ocr> (metadati specifici relativi al riconoscimento ottico del testo), <doc> (metadati specifici per oggetti digitali in formato testo che possono essere derivati o *born digital*) e <dis> (metadati specifici per la distribuzione di oggetti digitali): a seconda del tipo di oggetto digitale da descrivere e delle esigenze del singolo progetto queste sezioni possono essere più o meno necessarie ed essenziali.

Dal 2005, l'Istituto Centrale per gli Archivi - ICAR ha deciso di adottare MAG per i progetti di digitalizzazione di documenti archivistici⁵³ sviluppati nell'ambito del Sistema Informativo

(Biblioteca della Camera dei Deputati); Pierluigi Feliciati (Università degli studi di Macerata), Cristina Magliano, Patrizia Martini (ICCU); Maurizio Messina (Biblioteca Marciana di Venezia).

53 Cfr. Daniela Grana, *Le attività e i progetti di digitalizzazione nell'amministrazione archivistica*, in «DigItalia», numero 0 - 2005, pp. 92-96.

per gli Archivi di Stato – SIAS⁵⁴, per garantire la qualità effettiva dei progetti di digitalizzazione, contribuendo ad allargare l'adozione del profilo MAG in contesti extra bibliotecari e monitorandone l'utilizzo in contesti "misti" o che comunque includessero il trattamento di documentazione archivistica. Quindi, non si è trattato della scelta di un profilo di metadati per la gestione di depositi digitali archivistici accreditati, ma molto più semplicemente l'adozione di uno standard nazionale sperimentato per la gestione, la conservazione e la disseminazione via Web delle riproduzioni digitali di documenti d'archivio, garantendone soprattutto la stabilità delle relazioni con le descrizioni inventariali, non volendo sostituirle né con dei numeri né con i pochi elementi "orizzontali" offerti da Dublin Core⁵⁵.

Fin dalla pubblicazione del *Manuale* MAG nel 2006 si dava conto dello sviluppo di un "meccanismo per il quale MAG e METS possano interagire", visto il sostanziale parallelismo di approccio dei due standard strutturali per i metadati di gestione delle risorse in una biblioteca digitale. In un primo tempo "il Comitato MAG ha elaborato un applicativo (basato su un foglio di stile XSLT) in grado di trasformare un file MAG in un file METS, in modo da offrire agli utenti MAG la possibilità di usufruire di MAG per le proprie esigenze interne e, in generale, per coordinarsi a progetti che adottano il medesimo schema di codifica; allo stesso tempo però il sistema di conversione garantirà l'interscambiabilità e la condivisione dei dati a livello internazionale"⁵⁶. Recentemente il Comitato ha invece lavorato a un documento di *mapping* più elaborato, in corso di rilascio: in questo caso non si tratterà solo del confronto tra elementi informativi ma di un'analisi del modello di strutturazione di un documento MAG rispetto a uno METS, in rapporto al contesto di implementazione: non si dimentichi infatti che mentre MAG è stato sviluppato

54 Cfr. <http://www.archivi-sias.it/> e la pagina web dell'ICAR <http://www.icar.beniculturali.it/index.php?it/87/sias> .

55 Sul dettaglio delle motivazioni e sulle scelte tecniche attuate cfr. Pierluigi Feliciati, *Dalla descrizione archivistica al documento digitale: l'adozione del profilo MAG per la gestione della digitalizzazione negli archivi storici*, in «DigItalia», numero 1 – 2007, pp. 35-48. I dubbi espressi da alcuni colleghi archivisti sulla scelta di MAG sono perlopiù dovuti ad un fraintendimento, che ha portato all'errata convinzione che l'adozione di MAG per SIAS riguardasse il "dominio della descrizione e conservazione di archivi", addirittura per restituirne la "complessa struttura gerarchica", che certo - se così fosse - costituirebbe ovviamente "un tentativo, da alcuni considerato di limitato rilievo e di incerto risultato": cfr. Paola Carucci – Maria Guercio, *Manuale di Archivistica*, Carocci editore, Roma 2008, pp. 302-303.

56 *Manuale* MAG, cit., par. 1.4.3. Si consideri peraltro che in ogni caso MAG può essere usato come estensione di METS: si veda, per esempio, l'*application profile* realizzato dalla Biblioteca Provinciale di Campobasso in <http://web-serv.provincia.campobasso.it/biblioteca/digitale/>. Un altro esempio di uso parallelo di METS, MODS e MAG è rappresentato dal progetto BibIt- *Biblioteca Italiana*, una biblioteca digitale di testi rappresentativi della tradizione culturale e letteraria italiana dal Medioevo al Novecento, promossa dal "Centro interuniversitario Biblioteca italiana telematica" (CiBit), e gestita dalla sua unità attiva presso l'Università di Roma "La Sapienza". Cfr. http://www.bibliotecaitaliana.it/doc/docu_meta.php .

soprattutto nell'ottica di gestire risorse digitalizzate, METS è nato come contenitore per le risorse digitali e pertanto non definisce a priori il tipo di standard usato per descrivere le caratteristiche delle diverse risorse. In base all'uso che se ne fa, comunque, sia un documento MAG che uno METS possono essere usati come SIP, AIP, o DIP in un Open Archive System.

Venendo, per concludere, ai cosiddetti "metadati di conservazione", nel giugno del 2003 il Research Library Group dell'OCLC (*Online Computer Library Center*), lo stesso che ha redatto il volume *Trusted Digital Repositories: Attributes and Responsibilities*" (TDR)⁵⁷ ha creato un gruppo di lavoro internazionale sulle strategie di implementazione dei metadati di conservazione, il *Preservation Metadata: Implementation Strategies working group* (PREMIS WG)⁵⁸, composto da 30 esperti, che rappresentavano le biblioteche, i musei, gli archivi, le agenzie di governo ed il settore privato di cinque paesi diversi. L'obiettivo era quello di definire un insieme di base di metadati di conservazione e i risultati sono stati raccolti in un rapporto conclusivo, che contiene un modello per i metadati di conservazione (*PREMIS data model*)⁵⁹, cui è seguito un dizionario degli elementi semantici (*Data dictionary*), pubblicato nella sua versione 2.0 nel marzo 2008⁶⁰.

Che differenza c'è tra i modelli di metadati amministrativi che abbiamo sinora sinteticamente elencato e il modello PREMIS? Innanzitutto in PREMIS vengono incluse strutturalmente due variabili fondamentali che entrano in gioco per la conservazione digitale: la dinamica degli eventi nel tempo e i diversi agenti – umani o software – coinvolti. Le cinque maggiori aree rilevanti da coprire con i metadati di conservazione, secondo il PREMIS WG, sono infatti la provenienza (cioè le informazioni storiche sulla custodia dell'oggetto digitale, dalla sua creazione, ogni successivo cambio di custodia fisica e/o di proprietà), l'autenticità, l'attività di conservazione (ivi compresa qualsiasi conseguenza di azioni che impattino su forma, percezione o funzionalità dell'oggetto), l'ambiente tecnologico e la gestione dei diritti. PREMIS, inoltre, non formalizza un *application profile* nella forma che si è finora detta, ma definisce un modello composto di entità, relazioni e proprietà. Sono considerate entità un'entità intellettuale (un insieme coerente e unitario di contenuto), un oggetto digitale (un'unità discreta di informazione in forma digitale, che sia

57 <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>.

58 Sul PREMIS Working Group cfr. <http://www.oclc.org/research/projects/pmwg/>.

59 <http://www.loc.gov/standards/premis/>.

60 *Data Dictionary for Preservation Metadata: Final Report*, <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>. Si vedano inoltre le *Guidelines for using PREMIS with METS*, in: <http://www.loc.gov/premis/guidelines-premismets.html>.

un file, un *bitstream* o una rappresentazione, cioè una “materializzazione” digitale di un’entità intellettuale), un evento (qualunque azione nota occorsa al deposito digitale), un agente (una persona, un’organizzazione, o un programma software associato agli eventi di conservazione nella vita di un oggetto) e infine le asserzioni sui diritti.

Le entità sono ovviamente connesse da specifiche relazioni (vale a dire da associazioni tra le istanze delle entità nel tempo), che possono essere strutturali, di derivazione o di dipendenza. Infine, le proprietà delle entità (ecco i metadati!) in PREMIS vengono definite *unità semantiche*, sempre per non forzarle in uno schema, che possono essere anche contenitori che raggruppano un insieme di unità semantiche correlate. Il *data dictionary* prodotto dal gruppo di lavoro PREMIS è appunto un dizionario che comprende tutte le unità semantiche pertinenti che descrivono le quattro entità ed esprime le relazioni come informazioni di collegamento, includendo nelle informazioni dell’entità A un puntatore all’entità B correlata.

Il modello e il dizionario PREMIS sono il risultato di un processo di costruzione di consenso internazionale e multidisciplinare e possono aumentare le opportunità di applicazione in diversi tipi di istituzioni, di scenari della conservazione digitale e di sistemi di implementazione. Come schema di metadati di conservazione, se lo si vuole leggere come tale, tenta di rispondere a necessità pratiche in un’ottica di interoperabilità, vale a dire a documentare gli oggetti digitali nel tempo, rendendoli accessibili a lungo termine. Come si può ben immaginare, però, non è facile dirsi certi che l’efficacia del *data dictionary* PREMIS riuscirà davvero ad essere così ampia e soprattutto durevole.

Attualmente, mentre oltre che naturalmente in ambito OCLC e Library of Congress⁶¹ anche in Scozia⁶², in Germania⁶³, in Svezia⁶⁴, in Australia⁶⁵ e Nuova Zelanda⁶⁶ sono state avviati

61 Cfr. il *PREMIS Implementation Registry*, in <http://www.loc.gov/standards/premis/premis-registry.php>.

62 Il *Digital Data Archive (DDA) Project* dei National Archives of Scotland sta lavorando ad un sistema di ingest degli archivi governativi basato su OAIS e PREMIS. Cfr. <http://www.nas.gov.uk/recordKeeping/informationManagementPolicy.asp>.

63 Oltre alla collaborazione tedesca ad alcuni progetti internazionali che adottano PREMIS, va segnalato che il formato di ingest per l’OAI-AIP del *Federal Archives Digital Archive*, XBARC, si basa in parte su PREMIS. Cfr. <http://www.eha.de/wiki/index.php/Germany>.

64 Cfr. il progetto ARKIS II - *Swedish Archival Information System* (http://nad.ra.se/search_arch.aspx), che applica PREMIS per gestire le relazioni tra i metadati tecnici (objects), gli utenti (agents) e gli eventi in un sistema informativo di unità archivistiche in versione digitale. Si veda anche Göran Kristiansson, *ARKIS II - a Swedish Archival Information System*, in <http://www.knaw.nl/ecpa/sepia/workinggroups/wp5/1.html>.

65 Cfr. la *Australian Partnership for Sustainable Repositories* “...to establish a centre of excellence for the management of digital collections” (<http://www.apsr.edu.au/>) e in particolare al progetto PRESTA - *PREMIS Requirement Statement* (partner The National Library of Australia), i cui risultati sono consultabili in <http://www.apsr.edu.au/publications/presta/>.

66 Mi riferisco al progetto di *Statistics New Zealand Data Archive* (<http://www.stats.govt.nz/products-and-services/data-archive.htm>).

interessanti progetti di applicazione di PREMIS (vale a dire, di adattamenti del modello alle esigenze di specifici contesti applicativi), in Italia viviamo una fase di approfondimento, per cui vanno segnalati soprattutto l'importante iniziativa della Fondazione Rinascimento Digitale di Firenze, che ha organizzato un tutorial e un workshop su PREMIS lo scorso febbraio⁶⁷ e l'attività di stesura dei documenti di mappatura - dialogo tra MAG e PREMIS per la conservazione a lungo termine dei depositi digitali basati su MAG⁶⁸.

Venendo alle conclusioni, in Italia si discute spesso su quali schemi/profili applicativi di metadati e su quale modello di metadati per la conservazione adottare, purché il tutto sia ovviamente OAIS-compliant. Se la letteratura più recente sembra assegnare il ruolo di punta di diamante in questo campo allo schema di trasmissione METS, associato alla logica di conservazione permanente prevista nel *data model* di PREMIS, di certo questa accoppiata statunitense non sembra ancora in grado di risolvere tutti i problemi, specie per i contesti archivistici, anche se fornisce indubbiamente un solido e fertile terreno su cui far crescere soluzioni applicative che provino a rispondere alle complesse esigenze dei nostri contesti organizzativi e normativi.

Si tenga conto, poi, di una certa tendenza nell'*information science* internazionale ad investire molte energie per la creazione di nuovi standard internazionali o di nuove versioni di quelli già (talvolta faticosamente) digeriti dalle comunità nazionali (a loro volta particolarmente prolifiche di standard, specie in area anglosassone); questo fenomeno, specie se si vuol ragionare dell'utilizzo di tecnologie digitali nelle varie fasi di vita di un archivio, evoca una sorta di convergenza parallela tra elaborazione teorica e concreti profili applicativi.

Tra le tante criticità che l'adozione dei modelli provenienti dalla comunità *Library and Information Science* statunitense alle esigenze di produzione, gestione e conservazione di

67 Più precisamente, la Fondazione, insieme al Ministero per i Beni e le Attività Culturali e alla *Library of Congress* hanno organizzato a Roma, a febbraio 2009, due giorni di Workshop, composto da: un Tutorial e da una Tavola Rotonda. Il programma delle due giornate e tutti gli interventi in <http://rinascimento-digitale.net/PREMIS-workshop.phtml>. Nelle conclusioni del workshop (http://www.rinascimento-digitale.it/allegato_Rapporto%20in%20lingua%20italiana_7_631.phtml), si legge peraltro che “visto il grande interesse suscitato dall'applicazione dello standard PREMIS, il Ministero per i Beni e le Attività Culturali, la Fondazione Rinascimento Digitale e numerosi partecipanti hanno accolto con favore la proposta di creare un gruppo di lavoro internazionale, con la collaborazione della PREMIS Editorial Committee, per l'implementazione dello standard e il reciproco scambio di best practices tra i diversi repositories”.

68 I documenti di mapping sono in corso di redazione, ma per un resoconto su questo fronte di attività del Comitato MAG e per i punti critici dell'implementazione di PREMIS per un deposito digitale MAG cfr. l'intervento di Cristina Magliano al workshop PREMIS citato, in: <http://www.rinascimento-digitale.it/eventi/premis/premis->

complessi archivi digitali nell'intricato contesto amministrativo e normativo italiano⁶⁹, oltre alla necessità di far convivere profili di metadati pensati per depositi digitali “orizzontali” garantendo al tempo stesso le condizioni di autenticità (ovvero identità e integrità) dei documenti⁷⁰, uno dei problemi principali, mi pare, consiste nel declinare l'approccio canonico della rappresentazione archivistica, dal generale al particolare, usando i profili applicativi fin qui elencati, che costringono quasi sempre all'individuazione e implementazione di singoli “oggetti”.

Non è questa la sede per entrare analiticamente nel merito delle soluzioni adottabili: si vuole però osservare che una comunità archivistica dotata di punte teoriche avanzate come quella italiana, rinforzata anche dal confronto con il nostro incerto contesto normativo e politico-amministrativo e da certe criticità nel fare rete che ci contraddistinguono, ha tutti i numeri per dire la sua. Credo insomma si possa immaginare di mettere insieme – un po' sul modello australiano - le esperienze concrete e i gruppi di ricerca più avanzati per lavorare a un modello (un meta-profilo applicativo?) nazionale, unendo la ricchezza delle nostre competenze giuridiche e amministrative con la consapevolezza critica che abbiamo sui modelli di metadati standard utilizzabili per tutte le fasi di vita dell'archivio, per garantire qualità e stabilità nel tempo del documento digitale e del suo contesto, per offrire agli utenti modalità di accesso sicure ed avanzate.

Non si arriverebbe, forse, a sviluppare nuovi standard ma di certo sarebbe prezioso raccogliere quanto di meglio e di più opportuno c'è già per renderlo contestualizzabile attraverso la definizione di uno o più profili applicativi italiani di metadati per la gestione e conservazione a lungo termine degli archivi digitali (in senso proprio).

[workshop/Magliano-Premis.pdf](#).

69 L'evocativa definizione di “complex digital archives”, insieme ad altre giuste considerazioni sulle criticità dell'adozione di PREMIS per gli archivi, è di Mariella Guercio, nel suo intervento al workshop PREMIS citato: *PREMIS and the long-term preservation of complex digital archives. Lessons learned and critical issues from the CASPAR research*, in <http://rinascimento-digitale.net/eventi/premis/premis-workshop/Guercio-Premis.pdf>.

70 In proposito vedi soprattutto l'intervento di Giovanni Michetti, *Managing Authenticity in Digital Preservation*, al workshop PREMIS italiano citato, su cui cfr. <http://rinascimento-digitale.net/eventi/premis/premis-workshop/Michetti-Premis.pdf>.