

Appunti per la definizione di un set di *metadati* gestionali-amministrativi e strutturali per le risorse digitali <sup>1</sup>

---

<sup>1</sup> Bozza del 2001-01-30 / preparata da A. Scolari, M. Messina, C. Leombroni, G. Ciocchi, G. Bergamin per il *Gruppo di studio sugli standard e le applicazioni di metadati nei beni culturali* promosso dall'ICCU. Per gli obiettivi del gruppo vedi <http://www.iccu.sbn.it/grupmeta.htm>

|   |          |
|---|----------|
| <b>APPUNTI PER LA DEFINIZIONE DI UN SET DI METADATI GESTIONALI-AMMINISTRATIVI E STRUTTURALI PER LE RISORSE DIGITALI .....</b> | <b>1</b> |
| 1. INTRODUZIONE .....   | 3        |
| 2. IL MODELLO OAIS.....   | 4        |
| 2.1 Ambiente e componenti funzionali .....  | 5        |
| 2.2 Classificazione degli oggetti informativi.....  | 6        |
| 3. IL SET DI METADATI DELLA LIBRARY OF CONGRESS .....   | 7        |
| 3.1 Categorie funzionali dei metadati .....   | 7        |
| 3.2 Classificazione degli oggetti digitali.....   | 8        |
| 4. L'APPLICAZIONE BNCF DI UN SET DI METADATI GESTIONALE AMMINISTRATIVO E STRUTTURALE.....                                     | 9        |
| 4.1 La definizione degli archivi logici di metadati.....  | 9        |
| 4.2 Un workflow SIP -> AIP ->DIP .....  | 9        |
| 4.3 Un contributo alla classificazione.....   | 10       |
| 4.4 Scelta di metadati.....   | 10       |
| 5. CONCLUSIONI .....  | 11       |
| 6. APPENDICE (SET ARSBNI).....  | 12       |
| 7. BIBLIOGRAFIA .....   | 13       |

## 1. Introduzione

“Nel contesto degli oggetti digitali il termine *metadati* può essere associato a tre categorie funzionali:

1. *Descrittiva*: ha lo scopo di facilitare il recupero e l’identificazione della risorsa;
2. *Gestionale*: ha lo scopo di supportare la gestione della risorsa all’interno di una collezione;
3. *Strutturale*: ha lo scopo di collegare fra loro i componenti di oggetti informativi complessi<sup>2</sup>”

Il presente documento è dedicato alla seconda (*metadati gestionali-amministrativi*) e alla terza (*metadati strutturali*) categoria come definite dal rapporto appena citato. Gli oggetti digitali presi in conto sono sia quelli derivati da documenti preesistenti su supporti non digitali sia gli oggetti digitali nativi (i prodotti dell’editoria elettronica).

L’importanza dei *metadati* gestionali-amministrativi e strutturali diventa sempre più netta via via che emergono le problematiche poste sia dalla gestione “quotidiana” degli oggetti digitali (acquisire, gestire, archivi di file digitali), sia dalla necessità di garantire la cosiddetta conservazione digitale (cioè la conservazione e l’accesso a lungo termine degli oggetti digitali).

Di minore importanza è a prima vista la funzione di questo tipo di *metadati* ai fini della ricerca e del recupero della risorsa, tuttavia è insito nei *metadati* gestionali anche un aspetto di organizzazione degli oggetti digitali parte di una collezione (si pensi ad esempio alla sequenza delle immagini delle pagine di un libro digitalizzato), che diventa indispensabile per una corretta fruizione degli oggetti digitali stessi. Inoltre la presenza di questo tipo di *metadati* accanto agli oggetti digitali è da considerarsi necessaria e preliminare alla applicazione delle diverse possibili metodologie proposte per affrontare il problema della conservazione nel tempo.

Il presente documento viene inteso dal gruppo di lavoro come una prima indispensabile premessa all’obiettivo che ci si è posti. Infatti tenendo conto della esigenza di offrire con urgenza indicazioni circa l’utilizzazione di set standard di *metadati* gestionali ai numerosi progetti di digitalizzazione che sono in fase di avviamento in Italia, il gruppo ritiene suo obiettivo specifico quello di giungere in tempi per quanto possibili brevi alla predisposizione di un set minimo di *metadati* gestionali, che debbono diventare parte integrante di ogni progetto di digitalizzazione orientato alla produzione di

---

<sup>2</sup> *Preservation metadata for digital objects : a review of the state of the art : OCLC/RLG White Paper / authored by the Planning Committee of the OCLC/RLG Working Group on Preservation Metadata. - Draft for Working Group Review and Comment October 30, 2000. [non pubblicato, disponibile su richiesta a [giovanni.bergamin@bncf.firenze.sbn.it](mailto:giovanni.bergamin@bncf.firenze.sbn.it)]*

oggetti digitali derivati da documenti preesistenti su supporti non digitali o alla creazione di archivi di oggetti digitali nativi.

L'ambito naturale di applicazione di questo lavoro è quindi la BDI (Biblioteca Digitale Italiana) intesa come insieme dei progetti coerenti con determinate impostazioni. "Nonostante la frammentarietà delle esperienze di BD [Biblioteca Digitale] in Italia, non si può ignorare la presenza di episodi validi sotto ogni profilo. Tali esperienze hanno generato, all'interno delle strutture dove si sono realizzate, una serie di specifiche competenze che devono costituire il punto di partenza per qualsiasi futuro sviluppo nel campo<sup>3</sup>".

Nel seguito di questa relazione sono presentati il modello OAIS (Open Archival Information System), cui si sono conformate in modo esplicito o implicito le principali esperienze nazionali e internazionali nel settore, e due esperienze pratiche, che hanno operato scelte applicative di set di *metadati* gestionali, quella della *Library of Congress* e il progetto ARSBNi in corso presso la *Biblioteca Nazionale di Firenze*.

## 2. Il modello OAIS<sup>4</sup>

Per individuare i requisiti funzionali di un set di *metadati* gestionali-amministrativi e strutturali che abbia applicabilità generale è utile partire dalla definizione delle componenti chiave e dei processi su cui si basano le attività di archiviazione, gestione e conservazione di documenti elettronici, attraverso l'adozione di un modello di tipo logico-funzionale. La funzione svolta dal modello OAIS in questo contesto per i *metadati* gestionali-amministrativi e strutturali può essere paragonata alla funzione svolta da FRBR per i *metadati* relativi alla ricerca e al recupero delle risorse informative (ovvero la prima categoria funzionale definita all'inizio di questa relazione).

Il modello OAIS (Open Archival Information System) - ***Sistema Informativo Aperto per l'Archiviazione*** - adottato come modello di riferimento da alcune importanti esperienze del mondo di biblioteche ed archivi, quali NEDLIB e CEDARS in Europa, e PANDORA in Australia, intende appunto fornire questa cornice concettuale di riferimento: si tratta di uno standard ISO in fase di elaborazione, coordinato dal CCSDS (Consultative Committee for Space Data Systems) della NASA.

Benché sviluppato in origine dalla comunità della ricerca spaziale in relazione allo sviluppo di standard tesi alla conservazione a lungo termine dei dati ottenuti dall'osservazione dello spazio e del pianeta, il modello OAIS è adeguato allo scopo menzionato anche per altri tipi di comunità: infatti, pur essendo fortemente orientato al trattamento dei documenti elettronici, è applicabile a qualsiasi tipo di archivio - digitale o analogico; esso non specifica nessun tipo di implementazione e può essere applicato indifferentemente a oggetti digitali nativi, a prodotti di attività di digitalizzazione (quali files di immagini), e persino a oggetti fisici. Ciò che esso accentua fortemente è la tematica della *responsabilità* della conservazione; in modo particolare, della conservazione a lungo termine dell'informazione digitale.

L'OAIS si autodefinisce *"un'organizzazione di soggetti e sistemi che hanno accettato la responsabilità della conservazione dell'informazione e del mantenerla disponibile per una*

---

<sup>3</sup> Studio di fattibilità per la realizzazione della Biblioteca Digitale. Rapporto di sintesi. p.24

<sup>4</sup> Reference Model for an Open Archival Information System (OAIS) : CCSDS 650.0-R-1. Red Book. Issue 1. May 1999  
[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

*comunità determinata*"; si propone di fornire una solida base per una ulteriore standardizzazione nel contesto dell'archiviazione digitale e di promuovere una maggiore consapevolezza, da parte dei venditori di tecnologie, delle esigenze della conservazione a lungo termine dei documenti elettronici.

Riconoscendo che le collezioni digitali sono caratterizzate da una natura fortemente distribuita, e che vi è necessità, al tempo stesso, di sviluppare localmente politiche e procedure efficaci di gestione e conservazione di tali risorse, OAIS propone un modello di archiviazione distribuita, ma rispondente ad un ben individuato modello logico.

A tale scopo OAIS individua termini e concetti rilevanti per l'archiviazione di documenti digitali, identifica le componenti e i processi chiave comuni alla maggior parte delle attività di conservazione digitale, e propone un **modello logico di riferimento** per gli oggetti digitali e i *metadati* loro associati, che comprende la creazione e l'uso dei *metadati* utili a gestire il materiale elettronico, dalla fase di acquisizione a quella dell'accesso, fino alla conservazione.

Un'altra caratteristica del modello è che non esprime alcuna preferenza riguardo alle strategie di conservazione dei dati da adottare (p. es. migrazione o emulazione).

### 2.1 Ambiente e componenti funzionali

L'OAIS opera in un ambiente formato dall'interazione di quattro **entità**, definite come:

1. *i produttori (producer)*
2. *i consumatori (consumer)*
3. *il management*
4. *l'archivio stesso.*

Le **componenti funzionali** dell'OAIS sono definite come:

1. *Immissione (Ingest)*
2. *Archiviazione dei dati (Archival storage)*
3. *Gestione dei dati (Data management)*
4. *Accesso (Access)*
5. *Amministrazione dell'archivio<sup>5</sup> (Administration)*

A queste cinque componenti funzionali si aggiungerà, nella revisione del draft che è in corso, una sesta componente

6. *Pianificazione della conservazione<sup>6</sup> (Preservation planning)*

suggerita dal progetto NEDLIB (Networked European Deposit Library) nell'ambito della formulazione di un modello di DSEP (Deposit System for Electronic Publications), che specifica e rende "visibili" ulteriormente le attività di conservazione eseguite, a prescindere dalla strategia adottata.

<sup>5</sup> La differenza tra 3) *Gestione dei dati (Data management)* e 5) *Amministrazione dell'archivio (Administration)* è così convenzionalmente definita in OAIS: la prima componente coordina i *metadati* relativi all'informazione archiviata, la seconda ha in carico le operazioni giornaliere di mantenimento dell'archivio. Infine si noti la diversità tra la componente funzionale *Gestione dei dati (Data management)* e l'entità *Management*: quest'ultima è esterna all'archivio e si occupa delle politiche dell'archivio (cosa archiviare, come trovare i fondi ecc)

<sup>6</sup> C. Huc, L. Reich, D. Sawyer, *ISO Open Archival Information Systems Reference Model: concepts and current status* in <http://www.kb.nl/coop/nedlib/workshop/index.shtml>

## 2.2 Classificazione degli oggetti informativi

Oltre alle componenti funzionali, l'OAIS fornisce anche un **modello di strutturazione dei dati** adeguato a rappresentare l'informazione digitale da un punto di vista orientato alla conservazione.

Esso definisce un **Oggetto Informativo** (Information object) come composto da:

1. *Dati (Data)*
2. *Informazione sulla Rappresentazione (Representation Information)* necessaria per interpretare correttamente i dati.

In un ambiente digitale, i **Dati** (una sequenza di bits) sono collegati ad una **Informazione sulla Rappresentazione** che contiene tutto ciò che è necessario per rendere comprensibili quei bits, sia dal punto di vista strutturale (specificazione del formato, descrizione del s/w di accesso, etc.), sia semantico (p. es. in quale lingua è un testo in caratteri ASCII) .

Gli **Oggetti Informativi** possono comporre tre tipi di **I.P. - Information Packages**. Questi *Pacchetti di Informazione* sono dei contenitori concettuali di dati: ogni scambio di informazione da e per l'archivio e all'interno di OAIS, avviene attraverso l'utilizzazione di questi I.P.

1. **SIP - (Submission Information Package)** - *Pacchetto di Informazioni per l'Immissione*, utilizzato nella fase di immissione/acquisizione dei dati, mandato dal produttore in base al Submission Agreement stipulato con l'OAIS
2. **AIP - (Archival Information Package)** - *Pacchetto di Informazioni per l'Archiviazione*, destinato alla conservazione a lungo termine
3. **DIP - (Dissemination Information Package)** - *Pacchetto di Informazioni per la Distribuzione* trasferito dall'OAIS all'utente in base ad una richiesta di accesso.

Gli **Oggetti Informativi** possono appartenere a quattro classi:

1. *Informazione sul Contenuto(Content Information)*: l'oggetto primario trattato dall'archivio ( p. es. un documento in formato PDF, la documentazione del formato PDF)
2. *Informazione descrittiva per la conservazione(Preservation Description Information)*. Secondo OAIS vi sono quattro categorie di PDI
  - a) *Identificazione(Reference Information)*: enumera e descrive gli identificatori assegnati al Contenuto ( p. es. l'URN di **I**) *Informazione sul Contenuto* e il riferimento alla normativa sull'URN)
  - b) *Contesto(Context Information)*: documenta le relazioni del Contenuto con il suo ambiente (perché è stato creato, in che relazione è con altri Contenuti)
  - c) *Provenienza(Provenance Information)*: documenta la storia del Contenuto e i cambiamenti da esso subiti oltre che la catena di custodia (p. es. il formato originale dei dati, quale specifico processo ha permesso la creazione/trasformazione di un determinato oggetto informativo, chi è stato responsabile di quel processo)
  - d) *Autenticazione(Fixity information)*: documenta i meccanismi di autenticazione destinati ad assicurare l'integrità del Contenuto (p. es. l'impronta digitale calcolata con l'algoritmo MD5)
3. *Informazione sulla Composizione del Pacchetto di Informazione* ovvero *Informazione di Impacchettamento (Packaging Information)* (come reperire **1**) e **2**): p. es. in che disco, in che directory ecc.)
4. *Informazione Descrittiva (Descriptive Information)* (finalizzate alla ricerca e al recupero dell'informazione)

### 3. Il set di *metadati* della Library of Congress<sup>7</sup>

La LOC non dichiara esplicitamente di ispirarsi al modello OAIS per la strutturazione dei suoi archivi digitali, ma il set di *metadati* proposto conseguente alle esperienze fatte in vari progetti riconducibili al National Digital Library Program (NDLP), può sicuramente essere mappato in quel contesto.

Nell'ambito del presente documento l'esperienza della LOC appare significativa per i seguenti contributi:

- ✓ enumerazione puntuale di tutti i *metadati* utilizzati
- ✓ la precisazione delle categorie funzionali dei *metadati* e la classificazione degli oggetti digitali.
- ✓ il rapporto tra archivio bibliografico (*metadati* descrittivi) e archivio dei *metadati* gestionali e strutturali
- ✓ il set comprende *metadati* specifici anche per gli oggetti sonori ed audiovisivi, il che lo rende adatto ad essere impiegato all'interno di progetti di digitalizzazione di natura multimediale.
- ✓ il set della LOC può costituire un'utile base di lavoro - sia come quadro concettuale, sia come terminologia - per la proposta di un set standard per la BDI soprattutto per la parte relativa alle collezioni digitali create dalle biblioteche.

Qui di seguito si illustrano alcune caratteristiche di due aspetti considerati fondamentali.

#### 3.1 Categorie funzionali dei metadati

Anche la LOC distingue i *metadati* in base alla loro funzione:

- ✓ *descrittivi*
- ✓ *amministrativi*
- ✓ *strutturali*.

1. In generale si assume che i *metadati descrittivi*, destinati soprattutto al recupero degli oggetti digitali, siano registrati e gestiti al di fuori dell'archivio degli oggetti medesimi (ad es. a livello di OPAC) e con questi ultimi opportunamente collegati. Ciononostante alcuni *metadati* descrittivi possono in alcuni casi risultare utili nel flusso delle operazioni condotte all'interno dell'archivio digitale (es. nei passaggi fra gli "stati" SIP, AIP, DIP degli Oggetti Informativi, per usare la terminologia OAIS)
2. I *metadati gestionali-amministrativi* rendono eseguibili le funzioni di gestione degli oggetti digitali, in particolare quelle di presentazione all'utente e accesso da parte dell'utente, e quelle di conservazione (distinte in Digital Preservation: Conservazione digitale; e Digital Reformatting: Riformattazione)<sup>8</sup>.
3. Per quanto riguarda i *metadati strutturali* la LOC pone l'accento sulla loro funzione:
  - ✓ sia di indicatori della struttura delle relazioni fra gli oggetti digitali

<sup>7</sup>Table of core metadata elements, in <http://lcweb.loc.gov/standards/metadata.html>

<sup>8</sup>Preservation reformatting in ambiente digitale significa l'uso della digitalizzazione come alternativa al microfilm (si veda <http://lcweb.loc.gov/preserv/prd/presdig/presprinciple.html>). Digital preservation è la conservazione dell'oggetto digitale in quanto tale (Because the term "digital preservation" can have multiple meanings, including ones as contrasting as the preservation of digital data and the use of digital technology to preserve analog data, "life-cycle management" is intentionally used in the Preservation Digital Reformatting Program to refer specifically to the progressive technology and workflow requirements needed to ensure long-term sustainability of and accessibility to digital objects and/or metadata <http://lcweb.loc.gov/preserv/prd/presdig/preslifecycle.html>)

- ✓ es. 1: Parent\_Object\_ID: Identificatore di un oggetto correlato; usato per descrivere la struttura ad albero di oggetti complessi.
- ✓ es. 2: Relationship\_Type: Natura della relazione fra gli oggetti digitali, come lo status di singola pagina di un libro in rapporto al libro intero, etc
- ✓ sia di indicatori delle caratteristiche strutturali interne dell'oggetto digitale
  - ✓ es. 1: Audio\_Sampling\_Frequency: Numero di campioni per secondo di un oggetto sonoro.
  - ✓ es. 2: File\_Extension: Formato del file digitale, TIFF, GIF, MPEG etc.

### 3.2 Classificazione degli oggetti digitali

I diversi oggetti digitali sono organizzati in una struttura gerarchica a livelli. Si assume che ciascun oggetto digitale di un determinato livello abbia come attributi:

- ✓ i *metadati* specifici per gli oggetti digitali ascrivibili a quel livello
- ✓ i *metadati* specifici per gli oggetti digitali ascrivibili al livello immediatamente superiore.

La relazione fra *metadati* e livelli degli oggetti è espressa nel gruppo di colonne *Level* del *Table of Core Metadata Elements* della LOC. Un esempio di strutturazione gerarchica dei *metadati* è reperibile alla URL: [http://memory.loc.gov/ammem/techdocs/teams/samp\\_data.html](http://memory.loc.gov/ammem/techdocs/teams/samp_data.html)

I livelli degli oggetti digitali sono i seguenti:

- ✓ *Collezione digitale (Set)* : un insieme di oggetti digitali omogenei per contenuto, per titolarità (custodial responsibility) o altre caratteristiche salienti (es. il fondo musicale di Antonio Vivaldi della Biblioteca Nazionale Universitaria di Torino, ma anche le collezioni botaniche seicentesche delle biblioteche di Roma). Una Collezione digitale può identificarsi con un Aggregato (vedi oltre) o può essere composta di più aggregati;
- ✓ *Aggregato (Aggregate)*: un insieme di oggetti digitali omogenei per tipologia del contenuto e per titolarità (es. le partiture autografe di Vivaldi della BNT);
- ✓ *Oggetto primario (Primary Object)*: un oggetto digitale definibile come un intero coerente, in genere corrispondente ad un'unità fisica (es. un libro, un disco, o una singola partitura autografa di Vivaldi );
- ✓ *Oggetto intermedio (Intermediate Object)*: una particolare vista o formato dell'oggetto primario (es. un manoscritto in *formato immagine TIFF* o in *formato testo codificato in XML*), oppure una parte componente dell'oggetto primario (es. la digitalizzazione in real audio di *due facciate* di un disco a 78 giri ciascuna con vari brani, la *digitalizzazione* delle etichette);
- ✓ *Entità digitale (Terminal Object)*: il singolo file recante un'unità elementare di contenuto digitale (es. una singola pagina di una partitura autografa di Vivaldi, il frontespizio di un libro).



#### 4. L'applicazione BNCF di un set di *metadati* gestionale amministrativo e strutturale

Il modello OAIS è stato preso in conto sia nel progetto ARSBNI (digitalizzazione di frontespizi e indici della produzione libraria italiana degli ultimi quattro anni<sup>9</sup>), sia nei progetti che trattano i documenti nati digitali (NEDLIB<sup>10</sup> che si è chiuso il 31-12-2000 e EUROPE<sup>11</sup> la cui partenza è prevista nel secondo trimestre del corrente anno).

I contributi che vengono proposti alla discussione in questo contesto possono essere così sintetizzati :

- 1) La definizione degli archivi logici di *metadati* (bibliografico, oggetti digitali, *metadati*)
- 2) La definizione e l'implementazione di un workflow SIP-> AIP->DIP.
- 3) Un contributo alla classificazione degli oggetti digitali set /piece)
- 4) Una scelta di *metadati* (per la digitalizzazione di frontespizi e indici del libro moderno)

##### 4.1 La definizione degli archivi logici di *metadati*

L'implementazione ARSBNI ha distinto i seguenti archivi *logici*:

- √ **L'archivio bibliografico.** Si ritiene infatti che al di là di esigenze particolari o di pura sperimentazione l'accesso dell'utente agli oggetti digitali sia sempre mediato dall'archivio bibliografico (visto dall'utente come OPAC). In altre parole non si dovrebbe digitalizzare oggetti che non hanno "*metadati* descrittivi" già presenti in un archivio bibliografico standard.
- √ **L'archivio dei *metadati*** . Le informazioni necessarie a gestire l'acquisizione, l'archiviazione, l'accesso e l'amministrazione degli oggetti e delle entità digitali (in altre parole i *metadati* oggetto della presente relazione)
- √ **L'archivio dei dati digitali.** I file risultanti dal progetto di digitalizzazione.

ARSBNI ha scelto di non creare dei *metadati* descrittivi per gli oggetti digitali del progetto (considerando gli oggetti digitali prodotti alla stregua di nuove *Edizioni*).

In fase realizzativa l'archivio dei *metadati* e l'archivio dei dati digitali è stato disegnato per ospitare oltre che gli altri progetti di digitalizzazione anche tutte le acquisizioni di oggetti digitali nativi.

##### 4.2 Un workflow SIP -> AIP ->DIP

Nel progetto ARSBNI il SIP è composto da un insieme di file correlati.

Un file di tipo *xml*<sup>12</sup>:

- √ raccoglie tutte le informazioni sull'oggetto digitalizzato

<sup>9</sup> In questa sede è significativo aggiungere che ARSBNI sta recuperando su supporto digitale circa 5000 volumi del periodo 1957-1974 (mai pervenuti in BNCF o persi a causa dell'alluvione, ma presenti nell'OPAC della BNCF in quanto recuperati attraverso la retroconversione dei record trattati dalla BNI)

<sup>10</sup> <http://www.kn.nl/nedlib>

<sup>11</sup> <http://www.bncf.firenze.sbn.it/progetti/Europe/index.html>

<sup>12</sup> E' disponibile un esempio in <http://purl.oclc.org/net/sip>

- √ contiene la mappa di tutti i file relativi all'oggetto digitalizzato
- √ contiene anche informazioni bibliografiche e gestionali derivate dall'archivio bibliografico finalizzate a permettere il link tra l'archivio bibliografico e l'archivio dei *metadati*
- √ guida il caricamento e l'aggiornamento degli archivi descritti in 4.1

Il progetto prevede un workflow automatizzato: tutti i dati vengono raccolti per quanto possibile automaticamente (o tramite impostazioni di default)

L'AIP è il risultato della collaborazione di tutti e tre gli archivi descritti in 4.1

Il DIP viene creato su richiesta dell'utente grazie alla etichetta UNIMARC 956 (creata in fase di popolamento dell'archivio derivato OPAC) che permette il collegamento tra il record bibliografico relativo alla pubblicazione che è stata digitalizzata e l'archivio dei *metadati* gestionali. Grazie a quest'ultimo archivio le applicazioni sono in grado di presentare agli utenti le risorse digitali disponibili per quel record bibliografico.

### 4.3 Un contributo alla classificazione

L'archivio dei *metadati* distingue tra due tipologie di risorse digitali:

- √ una risorsa di tipo P (piece) che ha una corrispondenza uno a uno con una determinata entità digitale (Per *entità digitale* si intende uno specifico file - p. es. l'immagine digitale di una pagina).
- √ una risorsa di tipo S (set) che aggrega più risorse di tipo P e si riferisce ad un oggetto digitale (per *oggetto digitale* si intende un'insieme logico di file – p. es. tutte le immagini delle pagine che si riferiscono allo stesso libro).
- √ I set possono aggregarsi su  $n$  livelli gerarchici (p. es. il set relativo a un periodico digitalizzato, i (sub)set relativi alle annate, i (sub)set relativi ai fascicoli ...)

### 4.4 Scelta di metadati

Per la scelta dei *metadati* ARSBNi ha preso come base il set proposto da LOC.

L'appendice riporta la tabella dei *metadati* oggi usata dal progetto

## 5. Conclusioni

La definizione di un set di *metadati* amministrativi-gestionali e strutturali è un obiettivo non semplice. La complessità di un lavoro di questo tipo è sottolineata dalle conclusioni del *White paper dell'OCLC/RLG*<sup>13</sup>.

Accanto ai punti di convergenza :

- ✓ modello OAIS,
- ✓ l'indipendenza dello schema di *metadati* dalla tecnologia, dal tipo di oggetto digitale trattato, dal tipo di strategia di conservazione prescelta (migrazione, emulazione ecc).

Vi sono ancora dei punti di divergenza

- ✓ un set di *metadati* generali o specifico per ogni livello dell'oggetto digitale (collezione, oggetto digitale ecc)?
- ✓ a che livello di specificità si può ottenere un consenso sufficientemente esteso: un minimo di dati necessario per la conservazione dell'oggetto digitale e il resto lasciato alle implementazioni locali?

Su questo il Gruppo di lavoro (OCLC/RLG) prevede di lavorare almeno per tutto l'anno in corso. Per quanto riguarda il *Gruppo di studio sugli standard e le applicazioni di metadati nei beni culturali* promosso dall'ICCU il sottogruppo MAG ritiene che:

- ✓ vi siano oggi dei buoni punti di partenza (esperienze, studi, implementazioni) per quanto riguarda la definizione di un set di *metadati* gestionali-amministrativi e strutturali (uno di questi è certamente oggi lo *Studio di fattibilità per la realizzazione della Biblioteca Digitale*)
- ✓ quanto presentato nel presente documento sia solo un punto di partenza per la discussione
- ✓ il lavoro futuro del gruppo debba avere due direzioni: da un lato lo studio e la revisione di quanto prodotto a livello internazionale, dall'altro il confronto sul terreno dell'implementazione pratica con le implementazioni in corso o in fase di avvio
- ✓ essendo le esperienze della LC e del progetto ARSBNI dedicate primariamente ad applicazioni riferite a oggetti digitali derivati, sia necessario verificare la applicabilità del set di *metadati* che si elaborerà anche a oggetti digitali nativi
- ✓ sia inoltre necessaria una particolare attenzione affinché questo set possa risultare applicabile anche a progetti di digitalizzazione dedicati a materiali non librari

---

<sup>13</sup> *Preservation metadata for digital objects ... cit. pp. 34-35*

## 6. Appendice (set ARSBNI)

| <i>N</i> | <i>Nome del dato</i>       | <i>Note</i>   | <i>Origine preferita</i>   | <i>Scopo</i>  |
|----------|----------------------------|---|--|---|
| 0        | IDR                        | Identificatore univoco di oggetti/entità digitalizzati  | Attribuiti dal programma che popola l'archivio dati e <i>metadati</i>  | Identificare univocamente gli oggetti   |
| 1        | Sequenza                   | Numero d'ordine delle immagini  | Calcolata dal programma di digitalizzazione  | L'ordine progressivo delle entità che formano l'oggetto digitale: garantisce la corretta archiviazione delle informazioni e la possibilità di una corretta visualizzazione/navigazione degli oggetti digitali |
| 2        | Nomenclatura               | Nome dell'entità digitale (es. Frontespizio, Pagina 24 ...)   | Inserita di solito da chi digitalizza l'oggetto  | E' un informazione che completa la Sequenza: è il nome dell'oggetto digitale. In fase di visualizzazione permette ad esempio la scelta della pagina ecc.  |
| 3        | Nome del file digitale     | p. es. pag1.tif   | Calcolata di solito dal programma di digitalizzazione.   | E' il nome con il quale un s.o riconosce il file ☺  |
| 4        | Checksum                   | Il progetto usa l'MD5 .   | Può essere calcolata solo da un programma  | Serve ad assicurare l'integrità dei file nei trasferimenti e a verificare possibili duplicazioni nell'archiviazione delle entità digitali.. Garantisce anche un certo livello di autenticità del file.        |
| 5        | MIME Type                  | Tipi di file registrati su Internet (es application/pdf)  | Calcolata dal programma di digitalizzazione.   | E' lo standard con il quale i file vengono trasmessi e gestiti da un server web e da un browser   |
| 6        | Apparato di scansione      | Dati identificativi   | Calcolata dal programma di digitalizzazione.   | Informazione per la conservazione   |
| 7        | Software di Scansione      | Dati identificativi   | Calcolata dal programma di digitalizzazione.   | Informazione per la conservazione   |
| 8        | Parametri di scansione     | 8.1apparechiatura di scansione utilizzata<br>8.2 formato<br>8.3 compressione<br>8.4dimensioni<br>8.5 dpi<br>8.6 bit di profondità     | Calcolati dal programma di digitalizzazione sulla base delle decisione dell'operatore conseguenti agli standard di progetto. Molti di questi dati si trovano anche nell'header dei file prodotti | Informazione per la gestione (quale viewer usare in viaualizzazione) e alla conservazione (migrazione perché il formato non è più supportato)   |
| 9        | Data di scansione          | Formato ISO   | Calcolata dal programma di digitalizzazione.   | Informazione per la conservazione (per esempio in che data dovrò fare il refresh)   |
| 10       | Produttore della scansione | Chi ha la responsabilità del progetto   | Calcolata dal programma di digitalizzazione.   | Informazione per la conservazione   |
| 11       | Standard progetto          | Riferimento ad un documento esterno che precisa gli scopi del progetto e da conto sulle scelte effettuate (formati immagine, dpi ecc) | Inserita come default dal in progetto  | Informazione per la conservazione   |
| 12       | Nota pubblica              | Informazione supplementare che si desidera visualizzare con il file digitale  | Inserita dall'operatore  | Informazione destinata ad essere visualizzata con l'immagine  |
| 13       | Codice fruibilità          | Si tratta di un codice che rinvia ad una tabella esterna  | Inserita come default nel progetto ed eventualmente modificabile dall'operatore (es solo per utenti interni alla biblioteca, solo per utenti registrati ecc)                                     | Informazione per l'accesso  |
| 14       | Relazione                  | Codice di relazione con altri oggetti/entita (es. fa parte di / comprende)  | Calcolata dal programma di digitalizzaione   | Informazione per l'accesso  |
| 15       | Livello risorsa            | Codice che indica il posto occupato in una eventuale gerarchia: attualmente P(iece) o S(et)   | Calcolata dal programma di caricamento dei dati/ <i>metadati</i>   | Informazione per l'accesso  |

## **7. Bibliografia**

1. *Preservation metadata for digital objects : a review of the state of the art : OCLC/RLG White Paper / authored by the Planning Committee of the OCLC/RLG Working Group on Preservation Metadata.* - Draft for Working Group Review and Comment October 30, 2000. [non pubblicato, disponibile su richiesta a [giovanni.bergamin@bncf.firenze.sbn.it](mailto:giovanni.bergamin@bncf.firenze.sbn.it)]
2. *Studio di fattibilità per la realizzazione della Biblioteca Digitale / Intersistemi, Unisys per l'UCBLICE.* 2000
3. *Reference Model for an Open Archival Information System (OAIS) : CCSDS 650.0-R-1.* Red Book. Issue 1. May 1999 in [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)
4. C. Huc, L. Reich, D. Sawyer, *ISO Open Archival Information Systems Reference Model: concepts and current status* in <http://www.kb.nl/coop/nedlib/workshop/index.shtml>. 2000.
5. *Table of core metadata elements* in <http://lcweb.loc.gov/standards/metadata.html>. 2000.