




Definizione dei metadati necessari alla gestione della collezione di libri digitalizzati e dei relativi derivati per gli adempimento degli obblighi di cui all'art. 71-bis, commi 2-undecies e 2-duodeciesl (L. 633/641) conservate nel sistema di deposito digitale protetto del CSB

Editor	Sistema bibliotecario Sapienza - Angela Di Iorio, PhD
Contributor	Luca Lanzillo PhD, Agnese Galeffi Phd, Dott. Ezio Tarantino
Data stesura	Giugno-Luglio 2024
Revisione	Dicembre 2024
Destinatario	Responsabili del flusso di digitalizzazione libri accessibili Sapienza
Licenza	<p><u>Questa bozza è ad uso esclusivo di Sapienza</u>, la versione definitiva verrà rilasciata con la seguente licenza</p>  <p>Licenza Creative Commons Attribuzione - Condividi allo stesso modo 4.0 Internazionale</p>
Ultima modifica	15 Gennaio 2025

PREMESSA

Questo documento definisce i metadati necessari alla gestione della collezione di libri digitalizzati e dei relativi derivati per gli adempimento degli obblighi di cui all'art. 71-bis, commi 2-undecies e 2-duodeciesl (L. 633/641) conservate nel sistema di deposito digitale protetto del CSB.

Coerentemente con le politiche di cura del patrimonio digitale culturale, sviluppato dalle biblioteche Sapienza ¹, i libri digitalizzati ed i relativi derivati in formato accessibile verranno conservati nel sistema di conservazione della Sapienza Digital Library (SDL). Il sistema di conservazione è un sistema informatico (Rosetta ²), conforme allo standard ISO di conservazione Open Archival Information System (OAIS) ³, che mette Sapienza nelle condizioni di assolvere agli obblighi previsti per le "entità autorizzate".

La gestione dell'insieme delle misure di protezione delle opere sotto diritti e dei processi abilitanti alla loro fruizione in formato accessibile sono in via di implementazione, mediante l'integrazione delle funzionalità native del sistema di conservazione con quelle del catalogo Sapienza e con quelle

¹ Sistema bibliotecario Sapienza 2012-2020, <https://doi.org/10.13133/9788893771603>, pagg. 119-132

² Ex Libris Rosetta, https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation/940_Version_8.0

³ Giovanni Michetti,

http://www.interpares.org/display_file.cfm?doc=ip3_italy_dissemination_jar_michetti_digitalia_3_2008.pdf



di altri sistemi, al momento solo individuati, per la conversione automatica e la gestione di testi e materiale didattico in formato accessibile.

I libri digitalizzati ed i relativi formati accessibili, corredati di tutta la documentazione amministrativa comprovante la relazione giuridica tra il libro originale sotto diritti e la legittimazione a conservare i corrispondenti oggetti digitali. Una volta caricati nel sistema di conservazione di SDL gli oggetti digitali sono protetti da accessi non autorizzati e resi disponibili “unicamente ai beneficiari”, o eventualmente, “ad altre entità autorizzate”⁴ in base al modello di autorizzazioni del sistema, definito *ad hoc* per questa tipologia di oggetti.

Il sistema di conservazione SDL, conservando e proteggendo gli oggetti digitali nel lungo periodo mantiene i dati di gestione della catena di custodia, attraverso il mantenimento dell'integrità che abilita anche le garanzie di autenticità degli stessi. Inoltre, nativamente registra le operazioni effettuate sugli oggetti digitali, gestisce la relazione tra gli oggetti e pertanto, facilita sia la gestione, che l'accesso dei diversi formati. La gestione conservativa dei libri accessibili, infine riduce il rischio di operazioni ridondanti e abilita il riuso degli oggetti derivati con il conseguente miglioramento e velocizzazione delle procedure di accesso ai formati accessibili di opere già esistenti nel sistema. Di conseguenza nel lungo periodo, si ottiene un miglioramento complessivo del servizio agli utenti che hanno la necessità di fruire dei formati accessibili.

Analogamente, lo scambio con altre “entità autorizzate” viene facilitato sia per la registrazione delle informazioni di scambio, che per la gestione e l'aggiornamento delle informazioni previste per le categorie di beneficiari (comma 2 duodecies, art. 71-bis).

1. IL PROGETTO “I LIBRI SAPIENZA PARLANO” ED I LIBRI ACCESSIBILI

Il progetto “I libri Sapienza parlano” ha fornito alle persone con disabilità visive o DSA (tra le quali ipovedenti, non vedenti, dislessici etc.) la possibilità di consultare la stessa varietà di testi di cui possono usufruire coloro che disabilità non hanno. Il progetto si è sviluppato su due binari paralleli: la fornitura di libri letti della voce umana, esperienza più coinvolgente per alcuni testi, e di altri convertiti automaticamente e che garantiscono una maggiore versatilità nell'esplorazione del contenuto. La scelta di questo doppio output nasce dalla consapevolezza che le modalità di apprendimento, e dunque la possibilità di accedere alla cultura e all'informazione di ogni persona con disabilità, possono essere differenti a seconda della persona e del contesto di fruizione.

L'esperienza progettuale ha, inoltre, permesso di sviluppare un nuovo servizio per gli utenti di Sapienza, che a regime permette la digitalizzazione e la conversione in formato accessibile dei libri, per renderli fruibili alle persone con disabilità che ne hanno bisogno.

Di conseguenza la necessità di conservare in un deposito digitale l'intero corpus di libri digitalizzati, anche considerato il fatto che il servizio dei libri accessibili è destinato a durare nel tempo.

⁴ art 71-bis, comma 2-undecies

A questo scopo il deposito protetto della Sapienza Digital Library (SDL) presenta tutti i requisiti necessari per sopperire a questa esigenza e pertanto, di seguito vengono dettagliati gli elementi strutturali della SDL ed il modello di dati customizzato per questo tipo di collezione.

2. LA COLLEZIONE DIGITALE

Nella SDL la collezione digitale è una risorsa digitale costituita dai dati minimi di descrizione delle risorse digitali collezionate: dati descrittivi necessari ad identificarla e ricercarla nel sistema, nonché i dati che documentano il flusso di produzione e di provenienza della collezione.

La collezione digitale dei libri digitalizzati che rientrano nell'adempimento degli obblighi di cui all'art. 71-bis, commi 2-undecies e 2-duodeciesl (L. 633/641), verrà in primo luogo creata nel sistema di catalogazione Sapienza allo scopo di uniformare la procedura di acquisizione dei libri, che come requisito richiede l'esistenza dei dati descrittivi di catalogazione. Nel sistema di catalogazione, il codice identificativo della biblioteca creata per dare visibilità ai libri digitalizzati allo scopo, verrà utilizzato come identificativo della collezione digitale che raccoglie l'insieme di questa tipologia di libri, preceduto dal prefisso RMS. Ad esempio se il codice biblioteca è SB, il codice della corrispondente collezione digitale è RMSB.

Ognuna delle risorse raccolte nelle collezioni, manterrà il codice della biblioteca di provenienza del libro digitalizzato.

3. LA RISORSA DIGITALE A FORMATO ACCESSIBILE

Una risorsa digitale è costituita dall'insieme minimo e coerente di informazioni descrittive di un'entità intellettuale (PREMIS). Essa viene univocamente identificata e rappresentata come un'unità, nelle procedure di gestione della SDL (acquisizione dati e oggetti digitali e archiviazione). Una risorsa digitale è costituita, al minimo, dai suoi dati descrittivi e può essere un'unità intellettuale astratta, rappresentata dagli oggetti digitali di pertinenza o può essere un'aggregazione di altre risorse digitali.

Una risorsa digitale, creata e gestita dal CSB, comprende al minimo, i dati descrittivi e quelli di conservazione, definiti dal profilo di metadati della SDL.

La struttura concettuale della risorsa digitale è equiparata al Pacchetto Informativo (IP) dello standard di riferimento per la conservazione digitale, l'Open Archival Information System, e può essere utilizzata per le macro funzioni definite nello standard: Sottomissione (SIP), Archiviazione (AIP) e Disseminazione (DIP).

La sua struttura viene rappresentata come un insieme di dati conformi ai 1) metadati standard adottati, e definiti nel profilo di metadati, 2) i file come oggetti digitali di contenuto, che si



SISTEMA BIBLIOTECARIO
SAPIENZA

SAPIENZA
UNIVERSITÀ DI ROMA

strutturano in 3) modelli di contenuto utili a rappresentare in modo coerente le diverse tipologie di materiali culturali.

Una risorsa digitale rappresenta uno specifico contenuto intellettuale nel contesto della biblioteca digitale. Nei processi di gestione rappresenta l'unità informativa di riferimento.

Il modello di contenuto, specifico per i libri a formato accessibile, per la natura peculiare della tipologia di risorsa deve essere configurato in modo specifico per i metadati concernenti i diritti.

Ogni risorsa digitale è identificata, come in SDL dal codice biblioteca SBN preceduto dal prefisso RMS.

4. GLI OGGETTI DIGITALI DI FORMATO ACCESSIBILE

Nel contesto delle procedure di gestione e conservazione delle collezioni SDL, gli oggetti digitali sono quegli oggetti di contenuto che vengono prodotti dai progetti di digitalizzazione e che, necessariamente, appartengono ad una risorsa digitale specificamente identificata nel sistema e pertanto appartenente, al minimo, ad una collezione digitale. In questo caso la collezione sarà quella definita nel paragrafo n. 2.

Per oggetti di contenuto si intendono sia i file multimediali (oggetti digitali di contenuto), che rappresentano un contenuto intellettuale e sia quei file che contengono i dati - e relativi metadati - (oggetti digitali di metadati), che descrivono i file multimediali e la risorsa digitale di pertinenza.

In relazione allo standard di metadati per la conservazione digitale PREMIS, gli oggetti digitali SDL sono specificamente i file (di tipo contenuto o di tipo metadato) che vengono prodotti nelle diverse fasi della produzione della risorsa (digitalizzazione, deposito, archiviazione) che vengono raccolti nel deposito e successivamente trattati e trasformati in risorse digitali per essere archiviati e gestiti nel lungo termine.

5. ARCHIVIO DIGITALE PROTETTO DEPOSITO DIGITALE DI CONSERVAZIONE E ARCHIVIAZIONE (DEDICA)

L'archivio digitale DEDICA è un elemento infrastrutturale che raccoglie le collezioni digitali delle biblioteche Sapienza, nel contesto della SDL, il sistema informativo, che ne gestisce il suo incremento, il suo mantenimento e la sua conservazione e accesso nel lungo periodo.

L'archivio digitale DEDICA è uno storage di prossima consegna CINFO (entro autunno 2024) che verrà progressivamente utilizzato anche per le risorse gestite secondo quanto richiesto dal software Rosetta.



L'archivio digitale DEDICA svolge tre principali funzioni: 1) *deposito*, ovvero uno *storage* in cui vengono raccolti i materiali digitali secondo i criteri di ordinamento e produzione definiti nei progetti di digitalizzazione; 2) *archivio* destinato alla conservazione dove la gestione delle risorse digitali avviene attraverso procedure automatiche di servizio standardizzato; 3) *gestione dei dati*, ovvero l'insieme di basi di dati e di procedure software che gestiscono il ciclo di vita dell'archivio.

L'archivio è uno spazio di esclusiva gestione del Centro Sistema Bibliotecario (CSB) e viene gestito secondo procedure standardizzate per la razionalizzazione delle risorse e a minimizzare i rischi di perdita dell'informazione.

La funzione di gestione dei dati raccoglie, incrementa, arricchisce e mantiene i dati che descrivono le risorse digitali ed il loro ciclo di vita all'interno dell'archivio e del sistema informativo SDL. In altre parole, la gestione agisce sui dati utili alla conservazione e a mantenere l'accesso alle risorse nel lungo periodo. La gestione dei dati implica le funzioni di amministrazione della base di dati ed il mantenimento degli schemi locali e dei profili di metadati, della coerenza e della consistenza dei dati e della produzione degli oggetti digitali di metadati, in base alle procedure predefinite dal flusso di lavoro definito per la SDL.

6. METADATI PER I LIBRI CON FORMATI ACCESSIBILI

In coerenza con quanto sviluppato nel progetto Sapienza "I libri Sapienza parlano", i libri digitalizzati corredati dei formati accessibili, allo stato attuale possono avere due tipi di formato accessibile:

- formato immagine con testo OCR
- formato audio.

In base alla tipologia di formato il flusso di acquisizione SDL delle risorse, produrrà i metadati tecnici relativi agli oggetti (tipo di formato, dimensione file, date, checksum...).

Particolare attenzione deve essere posta all'aspetto dei vincoli legali che impediscono all'opera di essere digitalizzata (in quanto copia integrale), di diffusione (e quindi la pubblicazione) e quelli di riproduzione (ad esempio le copie per la fruizione come le copie di replica per la conservazione).

Lo stato dei vincoli legali che ostano alla riproduzione e diffusione dell'opera, deve essere verificato come prerequisito per la digitalizzazione/conservazione/accesso.

In questo caso gli accordi, le licenze, e i diritti di accesso, sono stati definiti, nell'ambito del progetto Sapienza.

Gli oggetti digitali prodotti devono essere replicati dal CSB solo a scopo conservativo ed il numero di copie da fare deve essere chiaramente definito da un documento formalmente approvato dal Consiglio del Sistema Bibliotecario.

Le copie devono essere firmate digitalmente e protette da password e conservate in uno storage a cui solo le persone autorizzate a fare il prelievo e la consegna all'utente, possono accedere.

Ogni accesso deve essere annotato in un registro digitale, con indicazione della data e dell'ora del prelievo/consegna e dell'identificativo della persona (in caso vi fossero più persone autorizzate) che lo ha eseguito.

Di seguito la configurazione specifica dei metadati concernenti i diritti.

Tipo di Diritto	Descrizione
Accesso	I materiali vengono solo conservati nel deposito. L'uso e la conservazione sono permessi solo agli aventi diritto.
Uso e riproduzione	Non Consentito per gli oggetti Consentito per i metadati
Copyright	Sotto diritti

I dati descrittivi delle risorse sono i dati provenienti dal catalogo e che vengono convertiti automaticamente nella struttura dati standard di SDL.

In conclusione, di seguito vengono elencati i metadati che sono stati definiti per la fase di digitalizzazione dei libri e che, nel corso della fase di acquisizione del deposito DEDICA vengono integrati con i dati descrittivi e di conservazione.

Nome metadato	Esempio	Obbligatorio
area disciplinare	Lettere	SI
codice biblioteca	RMSBV	SI
serie inventariale normalizzata		SI
inventario normalizzato		SI
appendice normalizzata		
barcode		SI
ocr	SI/NO	SI
audio	SI/NO	
titolo		
nome del file ocr		
n. pagine scansionate		
nome del file audio		
n. minuti registrati		
numero file audio		
lettrice/lettore		
testo desame	SI/NO	
note		