

Sistema di ricerca integrato: un nuovo catalogo di servizi per le biblioteche

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00011

Luigi Cerullo — Maria Cristina Mataloni

ICCU

Il Sistema di Ricerca Integrato (SRI) costituisce un articolato progetto di integrazione delle basi dati gestite dall'ICCU e al contempo mira ad offrire all'utente finale un servizio di restituzione originale, destinato a soddisfare le esigenze di un pubblico più vasto, costituito non solo da specialisti del settore.

All'interno di questo progetto se ne innestano altri complementari, SBNteca e SBNcloud, che hanno lo scopo di offrire alla comunità SBN strumenti tecnologicamente avanzati e di sicuro ausilio rispetto alla crescente domanda di servizi digitali, mai come oggi fondamentali per gli utenti finali.

Il modello di integrazione

L'idea di fondo da cui muove il progetto¹ nasce dall'esigenza di costruire un modello di integrazione dei dati di catalogo che consenta di superare le criticità legate alla presenza di piattaforme di ricerca e restituzione separate per ciascuno dei principali servizi informativi che l'ICCU garantisce (dal catalogo collettivo del Servizio Bibliotecario Nazionale alle banche dati specialistiche rappresentate dai censimenti nazionali, dall'aggregatore dedicato alle risorse digitali alle altre banche dati di corredo); al contempo si intende "ricucire lo strappo" determinatosi tra le informazioni catalografiche e le risorse digitali raccolte dall'aggregatore Internet Culturale.

Integrazione dell'informazione

In estrema sintesi l'integrazione è concepita come correlazione e sistematico riferimento a un oggetto informativo primario, rappresentato dal record bibliografico del Catalogo collettivo SBN, arricchito dal legame con la copia digitale presente nel sistema aggregatore Internet Culturale. Il dataset dell'Indice SBN, alla base del Catalogo collettivo, viene integrato sia da *collegamenti* a schede relative allo stes-

¹ Per un approfondimento di natura più tecnica, cfr. Luigi Cerullo, *ISS Project: The Integrated Search System in the National Bibliographic Services*, in: *Digital Libraries and Multimedia Archives. 14th Italian Research Conference on Digital Libraries, IRCDL 2018, Udine, Italy, January 25-26, 2018*, ed. by G. Serra, C. Tasso, Cham: Springer, 2018, (Communications in Computer and Information Science; 806), p 219-224.

so documento, presenti nel repertorio specialistico EDIT16² dedicato al libro italiano del 16° secolo, sia da schede di manoscritti provenienti dalla banca dati ManusOnLine³ (MOL), una tipologia documentaria non gestita nel Catalogo collettivo. In quest'ultimo caso ad essere recuperate sono le descrizioni interne dei manoscritti il cui profilo descrittivo meglio si concilia con il modello di restituzione dati del Catalogo collettivo.

Accanto all'integrazione dell'informazione riferibile ai "documenti", il progetto ha previsto lo sviluppo di strumenti e procedure finalizzate a costruire un *cluster* di record di authority per i nomi di persone ed enti collegati ai documenti. Anche in questo caso il modello di clusterizzazione fa perno sulla individuazione di una scheda informativa di riferimento, rappresentata parimenti dal record SBN che si arricchisce in tal modo di collegamenti ai repertori specialistici che continuano, dal canto loro, a conservare una propria rappresentazione della entità. Qualora l'entità non sia presente nel sistema di riferimento, la procedura si fa carico di integrare l'archivio Nomi di SBN con i nuovi record.

Una volta costituito, tale authority integrato continuerà ad essere alimentato attraverso l'ordinaria attività di catalogazione svolta nelle tre banche dati coinvolte (Indice SBN, EDIT16 e MOL), grazie alla "integrazione" del livello applicativo di cui si darà una breve panoramica nel paragrafo seguente.

Quanto descritto poc'anzi consente di raggiungere uno degli obiettivi primari di progetto: la possibilità di richiamare, direttamente dalla scheda di una entità (ente, persona o famiglia) - presente nel punto d'accesso unico per Nomi, esplorabile attraverso una innovativa interfaccia di ricerca guidata di cui si dirà nel prosieguo - l'insieme dei titoli riferibili a tale entità indipendentemente dalla loro provenienza, e di filtrare con semplicità i beni correlati scegliendo tra due relazioni primarie, semanticamente chiare, "documenti di" e "documenti su".

Integrazione applicativa

La sostenibilità di questo modello di integrazione dell'informazione è garantita dalla contestuale reingegnerizzazione dei gestionali EDIT16 e MOL, dotati della capacità di condividere i propri record - già in fase di creazione dell'informazione a cura degli operatori accreditati - con l'Indice SBN attraverso l'uso dei servizi rappresentati dal protocollo applicativo SBNMarc⁴.

² http://edit16.iccu.sbn.it/web_iccu/ihome.htm.

³ <https://manus.iccu.sbn.it/>.

⁴ <https://www.iccu.sbn.it/it/SBN/evoluzione-dellindice-sbn/protocollo-sbn-marc/index.html>.

I nuovi EDIT16 e MOL si configurano di fatto, nel nuovo assetto, come Poli speciali del Servizio Bibliotecario Nazionale, in grado di contribuire al Catalogo collettivo mantenendo le proprie specificità e autonomia.

Parallelamente allo sviluppo delle nuove applicazioni gestionali, la riorganizzazione del complesso sistema di aggregazione e fruizione delle risorse digitali dell'ICCU, rappresentato da Internet Culturale, ha permesso di sviluppare una nuova architettura dei flussi digitali articolata in tre linee di intervento - dettagliate nel prossimo paragrafo - che nel loro complesso consentono di raggiungere l'obiettivo di rappresentare il contenuto digitale come estensione del Catalogo, attraverso il sistematico riferimento delle "copie digitali" dei beni alle schede bibliografiche che li rappresentano.

Digital library distribuita

Internet Culturale continuerà a svolgere la sua funzione di infrastruttura dedicata alla raccolta e indicizzazione delle copie digitali rese disponibili da Teche remote, ma non disporrà più di un proprio portale di pubblicazione e di strumenti di ricerca e fruizione destinati all'utenza. Nell'architettura del nuovo sistema informativo il suo ruolo sarà quello di rendere disponibili al sistema di integrazione e indicizzazione centrale tutte quelle risorse digitali che siano riconducibili a una scheda catalografica già presente nel sistema di Catalogo, comunicando anche tutte quelle informazioni tecniche e d'uso che hanno la funzione di arricchire il *core* bibliografico e potenziare i suoi strumenti di ricerca sul patrimonio digitale.

Allo scopo di ampliare il ventaglio dei possibili fornitori esterni di contenuti digitali, l'aggregatore permetterà l'acquisizione anche dei soli metadati descrittivi, delegando alla Teca remota il compito di rendere visibile i propri contenuti digitali. Perché i metadati descrittivi siano facilmente allineabili alle schede catalografiche presenti in SBN e quindi nel Catalogo integrato del nuovo sistema informativo, l'Istituto pubblicherà nell'ambito di questo progetto una versione evoluta della Teca digitale italiana - strumento già in uso presso diversi provider di Internet Culturale - che consentirà, attraverso opportune mappature, la generazione automatica di tali metadati e la semplificazione dei processi necessari ad alimentare una propria teca digitale a partire dai dati di catalogo disponibili nei formati standard più diffusi.

Il nuovo Aggregatore digitale (questo è il nome tecnico corrispondente al suo rinnovato ruolo) si occuperà di gestire il "residuo" degli oggetti digitali non già associati agli elementi catalografici dei sistemi MOL, Editweb e SBNCloud, ovvero tutto ciò che i provider di oggetti digitali di SRI hanno al di fuori dei sistemi suddetti.

Allo stesso tempo disporrà di uno storage locale per gli oggetti digitali aggregati in modo da ovviare alle criticità dell'attuale sistema legate al reperimento dinamico degli oggetti digitali dalle Teche remote, consentendo in tal modo l'implementazione delle specifiche IIF⁵.

SBNTeca come Cloud centralizzato

Una delle linee progettuali chiave nell'economia complessiva del progetto è costituita dallo sviluppo di un sistema centralizzato di gestione del patrimonio digitale/digitalizzato. I servizi resi disponibili in tale sistema sarebbero orientati sia al recupero del "digitale sommerso", conservato dagli istituti presenti sul territorio, sia al trattamento di nuovi *asset* prodotti da campagne di digitalizzazione e/o produzione di contenuti digitali e multimediali.

La componente centrale di tale sistema è rappresentata da un DAM (Digital Asset Manager) i cui servizi saranno integrati negli ambienti gestionali specialistici di EDIT16 e MOL, e nella piattaforma SBNCloud destinata a potenziare il mondo della cooperazione SBN.

Tale componente opera dunque in modalità *multi-tenant* e come estensione funzionale pienamente integrata in tali ambienti gestionali, consentendo il recupero delle informazioni tecniche e gestionali sugli oggetti digitali direttamente nei contenuti catalografici.

Ecosistema informativo

La piattaforma di pubblicazione del nuovo sistema informativo è costituita da un network di Portali, ciascuno dei quali dà accesso a servizi di ricerca e tipologie di contenuti destinati a diverse comunità di utenti caratterizzate da esigenze informative diversificate.

EDIT16 e MOL disporranno di nuovi portali dotati di sistemi di ricerca ed esplorazione dei risultati basati su faccette e di una più razionale rappresentazione delle informazioni bibliografiche. Il caso di Manus è particolarmente emblematico del cambio di paradigma in questo senso, nella più articolata rappresentazione interna della scheda manoscritto che consente ora una più chiara individuazione delle unità testuali e del loro raggruppamento in unità codicologiche.

Grazie poi all'integrazione dei servizi del DAM della SBNTeca nei sistemi gestionali dei repertori specialistici, anche i loro portali disporranno della rappresentazione diretta

⁵ In termini di IIF Image API, il livello garantito è "2". Cfr. <<https://iif.io/api/image/3.0/>>.

dei contenuti digitali (digitalizzazioni parziali o integrali del documento fisico) attraverso il visualizzatore centrale dell'ecosistema, Mirador, basato sul protocollo IIIF.

Accanto alla reingegnerizzazione e rivisitazione funzionale delle piattaforme specialistiche di ricerca, il progetto ha previsto un analogo intervento sul portale del Catalogo del Servizio Bibliotecario Nazionale.

Il portale principale di progetto: Alfabetica

Grazie al core bibliografico rappresentato dal Catalogo generale e le sue logiche di integrazione, sommariamente descritte nei paragrafi che precedono, si intende superare il modello tradizionale di ricerca e restituzione bibliografica (quello di un OPAC, per quanto potenziato), in favore della costruzione di un modello di ricerca guidata da un rigoroso e tuttavia flessibile sistema di tassonomie.

Alfabetica – questo il nome del nuovo servizio – classifica tutte le risorse (i documenti) e le entità che nei cataloghi vi sono collegate (Nomi di enti, luoghi, persone) secondo le voci di una tassonomia doppia incrociata⁶.

Una tassonomia flessibile di questo tipo consente in primo luogo una classificazione multipla dei medesimi documenti grazie alla presenza di suddivisioni tipologiche ricorrenti in più categorie del sistema, ma allo stesso tempo considera una categoria tipologica (intesa come classificazione diretta di tipologie documentarie mappabili a priori con la classe) come “perimetro tematico” in grado quindi di richiamare tipologie documentarie, estranee come tali alla classe di interesse, in virtù del solo contenuto rappresentato. Individuare questo contenuto in assenza di chiavi semantiche esplicite (prodotto di una catalogazione semantica) è compito particolarmente complesso che ha comportato un lungo lavoro di analisi dei dati e poi di sperimentazione dei risultati di ricerca ottenuti in fase di implementazione del sistema di indicizzazione centrale.

SBNCloud

Il “framework di servizi” delineato nei paragrafi precedenti non si limiterà alla sola parte informativa, destinata al pubblico, ma si estenderà ai servizi gestionali attraverso lo sviluppo del nuovo ambiente integrato SBNCloud.

⁶ Utilizza un vocabolario controllato di parole per classificare i suoi oggetti, e lo organizza in due insiemi, l'insieme delle categorie (tassonomia verticale) e l'insieme dei tag (tassonomia orizzontale). Le categorie mappano i tipi di documenti con i canali di ricerca e rappresentazione dei contenuti del portale Alfabetica e identificano l'ambito di manifestazione primaria delle risorse; i tag (il contenuto informativo “parla di”), invece, portano il contenuto altrove classificando gli stessi beni (le loro manifestazioni secondarie basate sui tematismi individuati) con categorie e suddivisioni altre.



Figura 1.

SBNCloud costituisce la piattaforma di aggregazione dove saranno disponibili tutti i servizi e le informazioni relative al patrimonio di ogni biblioteca che voglia aderire a SBN usufruendo dei vantaggi, economici e gestionali, derivanti da un Cloud “software as a service” erogato direttamente dall’Istituto. Un radicale cambio di prospettiva che sposta l’enfasi sul servizio offerto e non più sul prodotto.

SBN: nuove esigenze crescono

La struttura stellare della rete SBN, con l’Indice al centro e i Poli in periferia, è stata pensata e mantenuta per rendere più agevole l’adesione a SBN da parte delle innumerevoli biblioteche di diversa tipologia diffuse su tutto il territorio italiano. Queste avrebbero così trovato nei poli, strutture preposte all’accoglienza in SBN, un presidio che si sarebbe occupato di erogare corsi di formazione, indispensabili per garantire una corretta adesione alla catalogazione partecipata, in grado di sobbarcarsi gli oneri dovuti alla gestione dell’hardware e software necessari al colloquio con l’Indice. Una struttura che avrebbe favorito la circolazione delle informazioni utili alla cooperazione, raggiungendo in maniera rapida e capillare le biblioteche collegate al Polo.

Questa organizzazione ha svolto egregiamente il proprio compito per molti anni, crescendo in numero (dal 1996 al 2020 i Poli sono passati da 33 a 104 e le biblioteche aderenti da 419 a 6.590)⁷.

⁷ Dati al 14 dicembre 2020. Per consultare il numero aggiornato di poli e biblioteche si veda la pagina <<https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche>>.

Da qualche tempo, tuttavia, la continua diminuzione del personale in servizio e i consistenti e costanti tagli agli enti pubblici hanno minato l'efficienza dei Poli nelle loro prerogative istituzionali, proprio nel momento in cui più forte si sentiva l'esigenza, da parte delle biblioteche, di trovare sostegno nella rete SBN. I tagli di cui si parlava colpiscono, infatti, indiscriminatamente tutte le strutture territoriali, dalle piccole biblioteche di quartiere alle pubbliche statali.

L'ICCU, come fornitore di servizi e coordinatore della rete SBN, sta da tempo riflettendo, insieme ai partner SBN, sulla necessità di un profondo rinnovamento e cambiamento del servizio stesso, che sia in grado di raccogliere le nuove esigenze e rispondere in maniera adeguata, fornendo alle strutture territoriali il giusto sostegno, cercando di sollevarle il più possibile da quelle attività tecniche indispensabili alla partecipazione alla rete SBN che al contempo sono costose in termini di risorse economiche e professionali. In tal modo soprattutto i Poli potrebbero orientare le proprie energie alle attività di informazione, formazione, indirizzo, controllo, indispensabili al buon funzionamento dei servizi e al miglioramento della qualità dei dati dell'Indice.

L'applicativo SBNWeb

Per la catalogazione partecipata e per la gestione dei servizi bibliotecari la comunità SBN ha a disposizione da anni, gratuitamente, l'applicativo SBNWeb, sviluppato dall'ICCU che ne garantisce tuttora la manutenzione.

L'Istituto ha recentemente pubblicato i codici sorgenti di SBNWeb e dell'OPAC di Polo sul *repository* di *code hosting* GitHub⁸, inaugurando così una nuova politica di distribuzione dei software che progetta e sviluppa per le biblioteche, gli istituti culturali e il mondo della cooperazione SBN. Il software è stato rilasciato con licenza *GNU AGPL v3*⁹ secondo quanto stabilito dalle *Linee guida su acquisizione e riuso di software per le pubbliche amministrazioni*¹⁰ dell'AGID¹¹.

Va detto, comunque, che SBNWeb è un applicativo andato in esercizio ormai più di 15 anni fa e la sua tecnologia risente del tempo trascorso. Inoltre, i continui adeguamenti alle esigenze di una eterogenea tipologia di biblioteche e la volontà di rendere molti servizi parametrizzabili hanno nel tempo reso il software sempre più complesso e, per le piccole realtà bibliotecarie, piuttosto sovradimensionato.

⁸ <https://github.com/IstitutoCentraleCatalogoUnicoBiblio?tab=repositories>.

⁹ <https://www.gnu.org/licenses/agpl-3.0.html>.

¹⁰ https://www.agid.gov.it/sites/default/files/repository_files/lg-acquisizione-e-riuso-software-per-pa-docs_pubblicata.pdf.

¹¹ Agenzia per l'Italia Digitale: <<https://www.agid.gov.it/it>>.

The screenshot shows the 'Sintetica biblioteca' (Synthetic library) interface. At the top, there is a navigation bar with 'SBN cloud', 'Menu', 'Polo sviluppo CSW', 'Seleziona biblioteca', and 'ITA'. On the right, it shows 'super user' and a 'Logout' button. Below the navigation bar, the title 'Sintetica biblioteca' is displayed. The main content is a table with the following columns: Polo, Codice SBN, Codice ISIL, Denominazione, Indirizzo, Tipologia biblioteca, Centro sistema, and Funzioni. The table contains four rows of data, each with a search icon in the Funzioni column.

Polo	Codice SBN	Codice ISIL	Denominazione	Indirizzo	Tipologia biblioteca	Centro sistema	Funzioni
AGR	24	IT-AG0078	Biblioteca della Provincia regionale di Agrigento Gaspare Ambrosini	piazzale Aldo Moro 1 92100 Agrigento	Specializzata		<input type="button" value="Q"/>
		IT-AG0134	Biblioteca del Polo territoriale universitario di Agrigento. Università degli studi di Palermo	Via Ugo La Malfa 92100 Agrigento	Istituto di insegnamento superiore		<input type="button" value="Associa"/> <input type="button" value="Q"/>
		IT-PA0671	Biblioteca del Polo didattico di Agrigento. Università degli studi di Palermo	Palermo	NON SPECIFICATA		<input type="button" value="Associa"/> <input type="button" value="Q"/>
CSW	01	IT-AG0005	Biblioteca dell'Archivio di Stato di Agrigento	Via Mazzini 185 92100 Agrigento	Specializzata		<input type="button" value="Associa"/> <input type="button" value="Q"/>

Figura 2.

Dal software al servizio

Il progetto SBNCloud nasce dall'esigenza di dare una risposta concreta alle criticità emerse nel tempo e brevemente descritte poc'anzi, sia per quanto riguarda l'aspetto amministrativo che le necessità tecnologiche.

Il nome stesso del progetto richiama il nuovo assetto su cui si baserà la prossima architettura, che verrà realizzata con software completamente open-source. Si tratta di una vera e propria rivoluzione rispetto alla concezione di servizio portata avanti dall'ICCU nei confronti degli utenti (sia interni che esterni) della rete SBN: offrire alle diversissime tipologie di biblioteche dei servizi scalabili in base alle altrettanto diverse esigenze cui tali servizi sono destinati.

Non più l'offerta di un applicativo monolitico, parametrizzabile sì, ma sempre complesso e ricchissimo uguale per tutti, bensì un servizio erogato in funzione di ciò che ciascuna entità SBN richiede effettivamente.

Quando si parla di entità ci si riferisce ai Poli tradizionali, alle singole biblioteche o anche ad aggregazioni di biblioteche (pensiamo alle tantissime reti territoriali, sistemi metropolitani, sistemi di comunità montane, già presenti e operanti all'interno di alcuni Poli).

In estrema sintesi, il progetto mira a:

- rendere meno onerosa sia in termini economici che operativi la gestione tecnica per Poli e biblioteche SBN. Soltanto le istituzioni di grandi dimensioni e meglio strutturate dispongono e disporranno delle risorse (non solo economiche ma innanzitutto tecnico-gestionali) necessarie a gestire in autonomia un sistema integrato orientato alla valorizzazione del patrimonio e all'erogazione dei servizi all'utenza. Le biblioteche di medie e piccole dimensioni rischiano l'emarginazione da

SBN proprio a causa dell'insostenibilità di costi e competenze;

- proporre tale architettura come servizio SaaS (Software as a Service) multi-tenant direttamente erogato dall'ICCU, in grado di ospitare anche singole biblioteche che trovano difficoltà a entrare in un Polo già costituito;
- dare maggior impulso a quei servizi divenuti via via più richiesti dall'utenza finale, come quelli relativi alla gestione del digitale in tutti i suoi aspetti, dagli strumenti di metadatazione alla gestione dei servizi di fruizione del digitale (DAM). Per quest'ultimo aspetto, SBNCloud utilizzerà il modulo apposito creato nell'ambito del progetto SRI (sistema SBNteca).

Il modello gestionale del progetto di sviluppo

La complessità del progetto ha richiesto l'adozione di un modello AGILE e di un approccio SCRUM¹² nello sviluppo delle componenti software dell'ecosistema che ha visto coinvolto sostanzialmente tutto il personale dell'Istituto. Il progetto del Sistema di ricerca integrato è stato suddiviso in dieci sotto-progetti (Work-package), ciascuno dei quali coordinato da un referente d'area per l'ICCU e da una figura corrispondente nella organizzazione del Raggruppamento temporaneo di imprese (RTI) aggiudicatario dell'appalto:

WP1 - Reingegnerizzazione di EDIT16 gestionale

WP2 - Reingegnerizzazione di MOL gestionale

WP3 - Sistema di importazione-come-riconoscimento (sviluppo del cluster d'autority)

WP4 - Reingegnerizzazione del backend di Internet Culturale

WP5 - Sviluppo del sistema SBNTeca

WP6 - Sviluppo del motore di ricerca integrato

WP7 - Sviluppo del Portale del Servizio Bibliotecario Nazionale (OPAC)

WP8 - Sviluppo del Portale di EDIT16

WP9 - Sviluppo del Portale di MOL

WP10 - Sviluppo del Portale principale: Alfabetica

Non meno complessa è l'articolazione del progetto di sviluppo complementare denominato SBNCloud e che vede coinvolto in primo luogo il personale dell'area SBN e dell'area Servizi.

I lavori si concluderanno nell'arco del 2021 con la progressiva pubblicazione di tutti i servizi tra la primavera e l'autunno.

¹² [https://it.wikipedia.org/wiki/Scrum_\(informatica\)](https://it.wikipedia.org/wiki/Scrum_(informatica)).

The Integrated Search System (ISS) represents a complex integration project of the databases managed by the ICCU and at the same time provide end user with an original research service, intended to meet the needs of a wider public, not only by sector specialists. Within this project there are sub-projects, the SBNTeca and SBNCloud, which aim to offer the SBN community technologically advanced tools, designed to provide a secure support in the face of growing demand for digital services.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020