

Il Polo digitale degli istituti culturali di Napoli

Luigi Cerullo

ICCU

1. Introduzione

L'idea di un polo di istituti culturali nasce nel luglio 2013, quando l'Istituto Italiano per gli Studi Storici¹, il Pio Monte della Misericordia², la Società Napoletana di Storia Patria³ e la Fondazione Biblioteca Benedetto Croce⁴ decidono di cooperare - nell'ambito delle misure adottate dalla Regione Campania e destinate alla digitalizzazione di archivi e biblioteche - unendo i fondi europei destinati all'acquisto delle attrezzature tecnologiche previste dal bando "Digitalizzazione e messa in rete di archivi e biblioteche pubblici e privati". Al nucleo originario si affianca nel dicembre dello stesso anno la Cappella del Tesoro di San Gennaro⁵.

1.1 I vantaggi della cooperazione

In primo luogo la condivisione delle risorse economiche, dovuta al contributo che ciascun istituto ha deciso di mettere a fattor comune, ha determinato la disponibilità di budget che ha consentito di progettare il sistema di gestione del patrimonio. La creazione del Polo ha reso possibile poi - sia grazie al prestigio delle istituzioni coinvolte sia grazie alla massa critica delle risorse potenziali - richiedere la collaborazione dell'ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche) nella fase di sviluppo di alcune delle componenti più innovative del software gestionale: in particolare la componente semantica che consente di generare dataset RDF a partire da fonti dati SBN (vedi paragrafi 6 e 7). Tale collaborazione è inoltre finalizzata al rilascio di una piattaforma open source, gratuitamente disponibile, che consentirà alle istituzioni della cultura di gestire l'intero ciclo di produzione e gestione del digitale finalizzato all'erogazione di servizi destinati alla comunità dei propri utenti.

¹ <http://www.iiss.it/>.

² <http://www.piomontedellamisericordia.it/home/listituzione/>.

³ <http://www.storiapatrianapoli.it/>.

⁴ <http://www.fondazionebenedettocroce.it/>.

⁵ <http://www.cappellasangenaro.it/>.

1.2 Il progetto di digitalizzazione

Il progetto è articolato in due ambiti che procedono paralleli: lo sviluppo della piattaforma di gestione (sulla quale è per lo più incentrato il presente lavoro) e l'attività di digitalizzazione del patrimonio, costituito prevalentemente da documenti d'archivio, da testi a stampa antichi e da materiale grafico.



Figura 1: un esempio del patrimonio documentale di pregio posseduto dagli istituti culturali, composto da manoscritti, disegni, materiale d'archivio, testi antichi e opere d'arte

I file prodotti dalla scansione ottica - prevalentemente attraverso l'uso di scanner planetari a sensore d'area - sono per lo più finalizzati ad offrire "copie per immagine" del bene; solo in taluni casi, tale processo è il primo passo verso la creazione di una vera edizione digitale dell'opera, attraverso l'estrazione automatica di testo ed entità significative e la conseguente marcatura basata sull'uso di grammatiche standard.

1.2.1 Il patrimonio oggetto di intervento in numeri

Circa 620.000 oggetti digitali nel formato master (considerando invece tutte le istanze prodotte si conteranno alla fine circa 1.800.000 file), così distribuiti in base al tipo di materiale:

- 330.000 pagine d'archivio di documentazione moderna e contemporanea
- 220.000 pagine di testi a stampa antichi
- 60.000 file relativi a manoscritti e pergamene
- 10.000 file di materiale grafico (fotografie, disegni e stampe).

1.3 Gli standard di digitalizzazione

I file vengono generati in formato TIF non compresso con campionamento a 300 ppi per tutti i documenti a stampa con dimensioni pari o superiori al formato A4; a 400 ppi per quelli inferiori a tale misura. Per i manoscritti, le pergamene e il materiale grafico (fotografie, stampe e disegni) il campionamento base sale a 400 ppi per arrivare fino ai 600 nel caso di documenti di dimensioni inferiori alla misura standard individuata. Quando le dimensioni o il particolare tipo di materiale lo richiedano, si prevede comunque di arrivare a valori di campionamento molto più elevati. Tutti i file sono prodotti con profondità colore di 24 bit RGB.

Tali file costituiscono la copia master per l'archiviazione e la conservazione. La conservazione degli oggetti è affidata a un servizio cloud di storage management che garantisce l'integrità dei file e consente, interfacciandosi con il sistema di gestione, di avviare servizi di delivery anche su tali formati.

I file derivati sono JPG con valori di campionamento che dipendono dall'uso: la piena risoluzione è destinata alla consultazione interna e alla vendita attraverso il sistema di e-commerce integrato nella piattaforma; i sotto-campionali (in media 150 ppi) saranno visibili on-line ed esposti tramite OAI-PMH⁶ verso gli aggregatori (vedi paragrafo 3.2).

2. Il sistema di gestione dei dati (Data Management)

L'architettura generale del sistema, da un punto di vista funzionale, è piuttosto complessa e prevede la presenza di numerosi moduli che interagiscono tra di loro al fine di erogare i servizi necessari a supportare i complessi processi di lavoro finalizzati alla descrizione, digitalizzazione, conservazione e divulgazione di beni culturali documentali e museali. Punto di forza della soluzione adottata è la disponibilità, in un ambiente integrato Web based, di tutti i servizi software dedicati in primo luogo alla descrizione del patrimonio e alla sua valorizzazione, poi, attraverso la gestione del ciclo di produzione e disseminazione in ambiente digitale.

2.1 I modelli dei dati

Il sistema di gestione dei dati prevede diversi moduli di catalogazione e descrizione dei materiali:

- Un modulo di descrizione archivistica basata sullo standard ISAD (G)⁷ e la sua codifica XML EAD (Encoded Archival Description)⁸ che rende le descrizioni prodotte con tali sistemi indipendenti dalla particolare architettura software implementata, consentendone agevolmente l'esportazione verso altri sistemi

⁶ <https://www.openarchives.org/pmh/>.

⁷ ISAD (G): General International Standard Archival Description, le cui specifiche sono disponibili all'indirizzo: <<http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>>.

⁸ <http://www.loc.gov/ead/>.

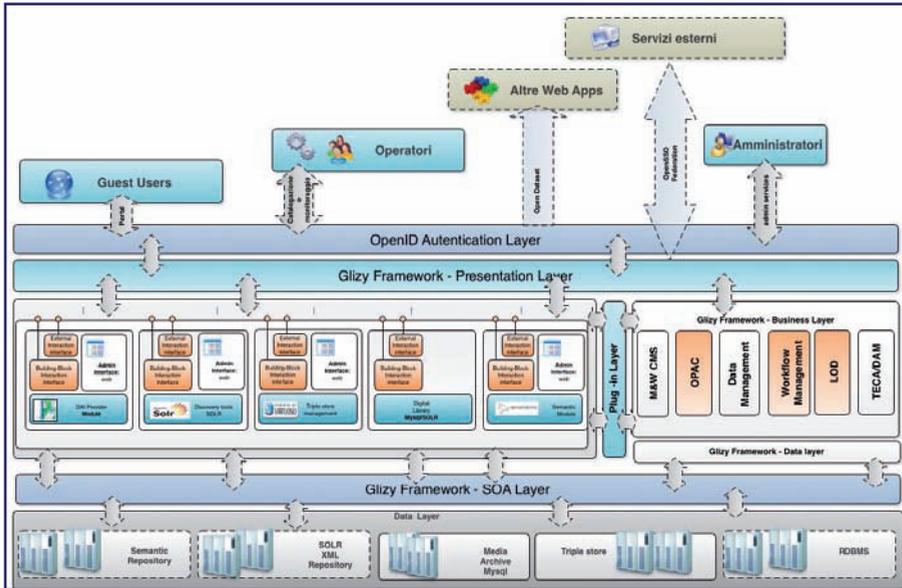


Figura 2: *diagramma funzionale dell'intera piattaforma di gestione*

che adottino tale standard per l'interscambio dei dati. In generale ogni modulo del sistema di gestione del Polo è basato su uno standard di struttura e di scambio dati che garantisce al catalogo creato la sua esportabilità su altre piattaforme. Il profilo applicativo studiato per il Polo consente la descrizione di complessi archivistici di documentazione moderna e contemporanea fino ad un livello analitico (ISAD 4, livello di descrizione dell'unità documentale).

In linea generale la catalogazione in Museo&Web FAD - sistema che rappresenta la base open-source nonché la componente funzionale core del sistema di gestione del Polo da cui si è partiti nello sviluppo, affidato al Gruppometa srl⁹, partner tecnico del presente progetto - permette di utilizzare un insieme di strumenti che agevolano il compito dell'operatore: elenco dei campi già compilati, gestione di liste e vocabolari (authority files), campi ripetibili, ecc. Tali strumenti sono forniti all'interno di interfacce grafiche di semplice utilizzo descritte di seguito.

- Un modulo che consente la catalogazione del patrimonio sulla base delle normative ICCD (Istituto centrale per il catalogo e la documentazione). In tale caso l'applicazione consente l'import-export nel formato standard definito dall'ICCD (trc 92). Le normative previste sono: F (fotografia)¹⁰, D (disegno)¹¹, S (stampa)¹² e OA (oggetto d'arte)¹³.

⁹ <http://www.gruppometa.it>.

¹⁰ <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/10>.

¹¹ <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/21>.

¹² <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/33>.

¹³ <http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/29>.

- Un modulo di integrazione con l'ambiente SBN (Servizio bibliotecario nazionale) basato su un sistema di cooperazione applicativa che consente al software del Polo di importare i dati bibliografici dal Polo SBN che fa capo alla Biblioteca nazionale di Napoli (tale Polo costituisce per i partner del progetto l'ambiente di lavoro SBN, rappresentato dal software gestionale Web-based SBNWeb¹⁴, sviluppato dall'ICCU e usato da oltre 20 poli SBN). Tale scelta è dettata dall'esigenza di non investire risorse nella duplicazione di servizi che sono già disponibili sul territorio e ampiamente testati, oltretutto dall'esigenza di rendere massima la visibilità del patrimonio librario e grafico grazie all'immediata disponibilità dei dati bibliografici resa possibile dal sistema informativo SBN.
- Un modulo che consente la descrizione dei manoscritti basata sullo standard TEI-MS (Text Encoding Initiative Manuscript Description)¹⁵ adottato dalla piattaforma Manus-online dell'ICCU, che permette di importare le descrizioni in tale sistema informativo che rappresenta di fatto il censimento dei manoscritti posseduti dalle biblioteche italiane.

2.1.1 Schema del modulo di gestione dati con tutti i canali gestionali esplicitati suddivisi nei principali domini descrittivi

1. Patrimoniale:
 - a. OA (opera d'arte)
 - b. F (fotografia)
 - c. D (disegno)
 - d. S (stampe)
2. Archivistico:
 - a. EAD (ISAD compliant): standard internazionale di codifica della descrizione archivistica
 - b. EAC-CPF¹⁶ (ISAAR CPF¹⁷ compliant): standard internazionale di codifica per la descrizione dei soggetti produttori
3. Bibliografico:
 - a. SBN Marc (in cooperazione applicativa)
 - b. MODS¹⁸ (Metadata Object Description Schema)
 - c. TEI-MS (Editor TEI Manus compliant)

¹⁴ <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/SbnWeb/>.

¹⁵ <http://www.tei-c.org/release/doc/tei-p5-doc/it/html/MS.html>.

¹⁶ Encoded Archival of Context, standard di struttura basato su metalinguaggio xml conforme allo standard di contenuto ISAAR: <<http://eac.staatsbibliothek-berlin.de/>>.

¹⁷ ISAAR (CPF) Standard internazionale per i record d'autorità archivistici di enti, persone, famiglie

¹⁸ <http://www.loc.gov/standards/mods/>.

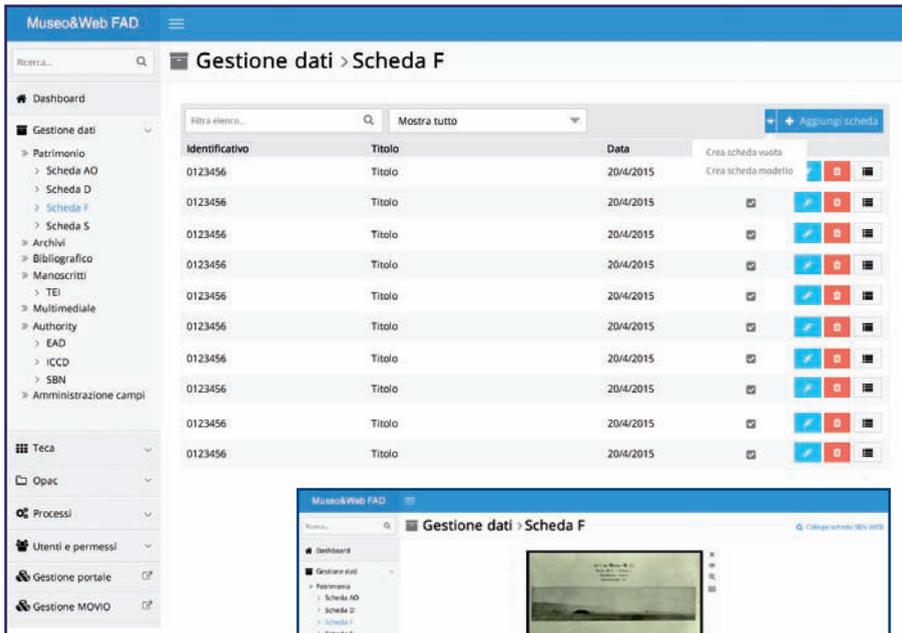


Figura 3: prototipo di GUI del sistema di data management con una esemplificazione della scheda ICCD F (fotografia)

2.2 Il sistema di cooperazione applicativa

Favorire l'integrazione del catalogo di biblioteca (la cui funzione precipua è identificare e localizzare una risorsa

documentale) con i cataloghi che descrivono il bene culturale sulla base di modelli specialistici è uno degli obiettivi di tale progetto: in altre parole favorire l'integrazione tra gli ambienti descrittivi, in particolar modo quando questi si sovrappongono parzialmente perché restituiscono, a livelli di approfondimento diversi e per utenze differenziate, le stesse entità.

È il caso dei disegni, delle stampe e delle fotografie per le quali il sistema del Polo digitale prevede l'integrazione tra l'ambiente SBN, finalizzato alla descrizione immediatamente disponibile e fruibile attraverso il sistema informativo nazionale rappresentato dall'OPAC SBN, e i cataloghi basati sulle normative definite, per tali beni, dall'ICCD.

Per il Polo digitale, SBN infatti è l'ambiente cui è riservata la fase di record-ma-

king bibliografico (che include anche materiale grafico come stampe, disegni, fotografie): il materiale librario viene gestito integralmente in ambiente SBN e importato – attraverso uno scarico dati in formato Unimarc – allo scopo di alimentare il sistema informativo del Polo digitale. Il materiale grafico (disegni, stampe e fotografie) viene, in tale scenario, preliminarmente catalogato in SBN, per poi essere dinamicamente importato, sulla base del protocollo applicativo su cui si basa il funzionamento di SBN (SBNMarc¹⁹), nel sistema di gestione del Polo digitale che prevede, per tali materiali, i tracciati definiti dall'ICCD (scheda F per la fotografia, scheda D per il disegno, scheda S per le stampe). Tale derivazione presuppone lo studio di mappature tra i modelli dati adottati nei due ambienti, che è tra gli scopi di una collaborazione scientifica avviata tra l'ICCU e il Polo digitale.

2.2.1 Descrizione funzionale

Gli istituti che partecipano al progetto catalogano, inventariano e collocano il materiale grafico, come detto, con le procedure di SBNWeb, usando le normali utenze disponibili sull'ILS²⁰ di Polo. I record così creati sono poi estratti da Indice o da Polo in cooperazione applicativa tramite protocollo SBNMarc – attraverso una interfaccia di ricerca dedicata, disponibile in Museo&Web FAD (componente di data management della piattaforma di gestione del Polo digitale) – con l'invocazione di un servizio specifico (“cerca”) tra quelli ammessi dal protocollo.

Le procedure cablate nel sistema d'arrivo, definite sulla base delle mappature studiate, consentono la precompilazione dinamica dei campi ICCD mappati con le strutture dati SBNMarc importate.

Lo scopo di tale processo è consentire, attraverso un unico flusso di lavoro, di alimentare sia il sistema informativo SBN che i repository ICCD locali, estendendo e arricchendo sul sistema di gestione del Polo digitale la descrizione dei suddetti materiali, attraverso i tracciati specialistici disponibili secondo il livello di descrizione scelto dai diversi istituti (inventariale, pre-catalogo e catalogo).

L'output finale – grazie ai formati di scambio standard previsti in fase di export dei dati – sarà un pacchetto generato in formato ICCD92 che potrà essere importato nelle banche dati regionali e centrali.

Nell'ambito di tale progetto è allo studio la fattibilità di una integrazione con il sistema SigecWeb²¹ in modo analogo a quanto previsto con l'ambiente SBN. Infatti per tale sistema sono documentate API di tipo Web services²², principalmente per

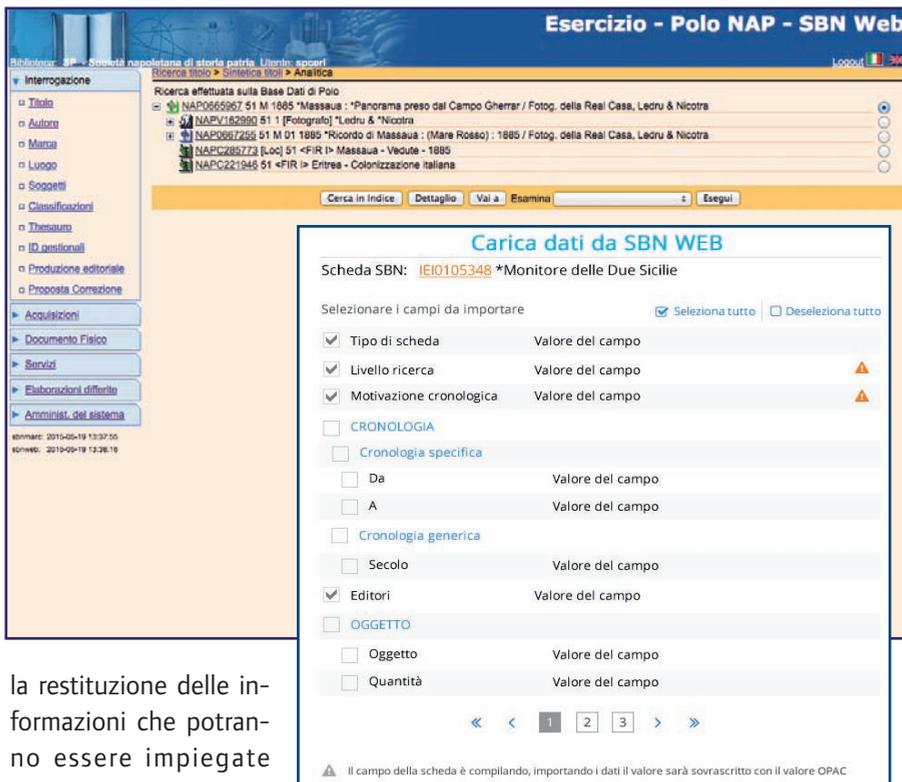
¹⁹ Le specifiche del protocollo alla base del sistema di cooperazione applicativa attraverso il quale è realizzato il catalogo collettivo sono disponibili all'indirizzo:

<http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/evoluz_indice_sbn/pagina_143.html>. Il protocollo (nella versione 2.0) consente anche la gestione del materiale audiovisivo.

²⁰ Integrated Library System, rappresentato da sbnweb.

²¹ <http://www.iccd.beniculturali.it/index.php?it/118/sistema-informativo-generale-del-catalogo-sigec>.

²² <http://www.iccd.beniculturali.it/index.php?it/394/interoperabilit>.



la restituzione delle informazioni che potranno essere impiegate per l'importazione puntuale di singoli record, mentre per l'importazione batch nel sistema locale di schede eventualmente già presenti nel SigecWeb si userà il formato di scambio ICCD92 sopra citato.

Figura 4: nella prima immagine è rappresentato l'ambiente di recordmaking SBN costituito dal gestionale di Polo SBNWeb. Il protocollo applicativo SBNMarc consente al Polo digitale di acquisire dinamicamente i dati creati in SBN popolando, sulla base delle mappature previste, i campi della normativa ICCD corrispondente

3. Digital Assets Management (DAM) e funzionalità di Teca digitale

La componente funzionale fondamentale in una piattaforma gestionale di tale natura è indubbiamente il sistema di digital library. Il progetto prevede lo sviluppo di una teca digitale che consentirà al personale di gestire i file digitali prodotti (che siano originati da campagne di scansione – come in questa prima fase – o che siano prodotti internamente sulla base di richieste di servizi – ad esempio richieste di digitalizzazione avanzate dagli utenti attraverso i canali di richiesta disponibili sul sito-portale), attraverso la creazione dei metadati gestionali. In ambiente digitale i metadati gestionali sono fondamentali per consentire, attraverso un sistema di interfacce informatiche, di ricostruire virtualmente l'integrità fisica e la struttura logica di un documento analogico e del testo/contenuto che esso veicola. Accanto

ai metadati tradizionalmente associati alla ricerca in un catalogo (valori che garantiscono le funzioni di ricerca, localizzazione e selezione di una risorsa informativa), risultano imprescindibili i metadati di natura strutturale che consentono la corretta rappresentazione della struttura fisica del documento analogico (l'organizzazione delle parti costituenti il supporto documentale) e della struttura logica dell'interfaccia testuale (ad esempio i capitoli di un libro). Tali metadati aggiuntivi vengono generati manualmente attraverso la disponibilità di editor sviluppati ad hoc.

Per garantire la massima flessibilità di utilizzo, nonché l'integrabilità con la componente di data management Museo&Web FAD, una facile manutenibilità e prestazioni adeguate, scalabilità, sicurezza e flessibilità d'uso, si è deciso di implementare una soluzione di Digital Asset Management (DAM) per la gestione degli oggetti digitali basato su framework Glizy scritto in PHP, motore di ricerca Apache SOLR e un front-end realizzato in AngularJS²³.

Caratteristica principale del DAM è la possibilità di specificare in modo estremamente accurato tutti i set di operazioni che possono essere effettuate su ogni singolo oggetto digitale come:

- inserimento, modifica e cancellazione degli oggetti digitali
- ricerca e consultazione degli oggetti complessi²⁴
- editing dei metadati nei formati supportati
- gestione delle collezioni o delle raccolte per facilitarne il riuso
- tagging gerarchico dei contenuti al fine di facilitare le operazioni di ricerca
- conversione degli oggetti digitali nel maggior numero di formati compatibili
- gestione di oggetti gerarchici
- gestione di oggetti multiformato
- supporto al protocollo OAI-PMH

3.1 Profili di metadati generati e workflow di metadattazione

Le linee dedicate alla codifica dei metadati gestionali contemplano processi che è possibile definire come semi-automatici. Una volta manipolato attraverso le funzionalità del DAM, l'oggetto digitale - che sia rappresentato da un singolo file o da una collezione di file - è agganciato, attraverso apposite interfacce di ricerca, alla fonte dati descrittiva dell'entità intellettuale rappresentata dal record bibliografico o dalla scheda archivistica. Tale fonte, grazie a mappature cablate nel sistema, consente il popolamento automatico della sezione descrittiva dei metadati gestionali.

²³ Framework javascript opensource, supportato da Google. Angularjs è una infrastruttura per la creazione di applicazioni client side. Dalla documentazione ufficiale: «... Angular è quello che HTML avrebbe dovuto essere se fosse stato progettato per sviluppare applicazioni ...». Per approfondimenti: <<https://angularjs.org/>>.

²⁴ Un oggetto complesso in tale ambiente è rappresentato da una associazione tra bytestream (il vero e proprio oggetto digitale) e datastream (metadati espliciti definiti in un qualsiasi schema).

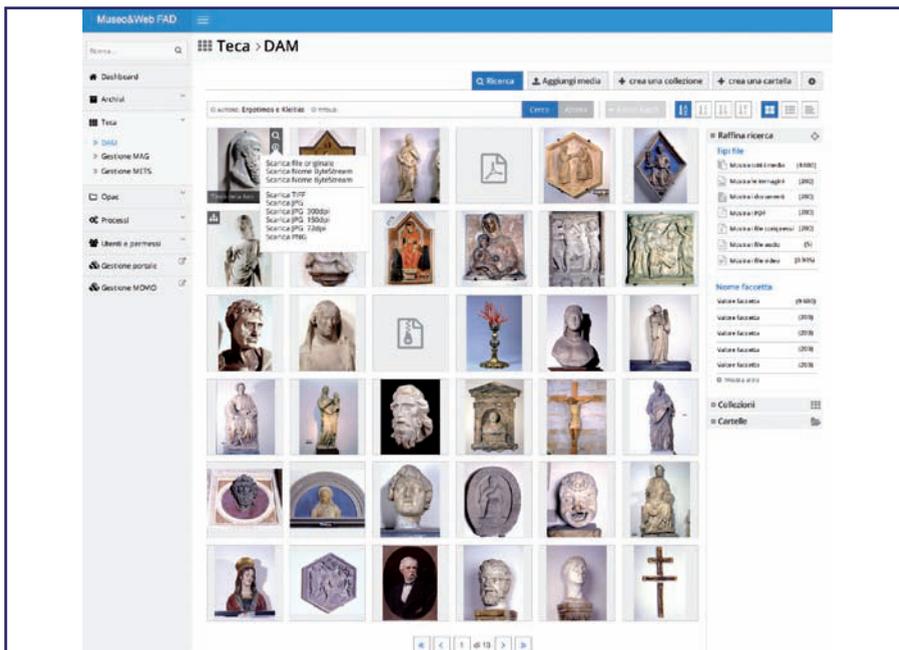


Figura 5: generazione dei formati derivati a partire dal master conservativo, operazione effettuata su singolo oggetto attraverso le funzionalità del DAM

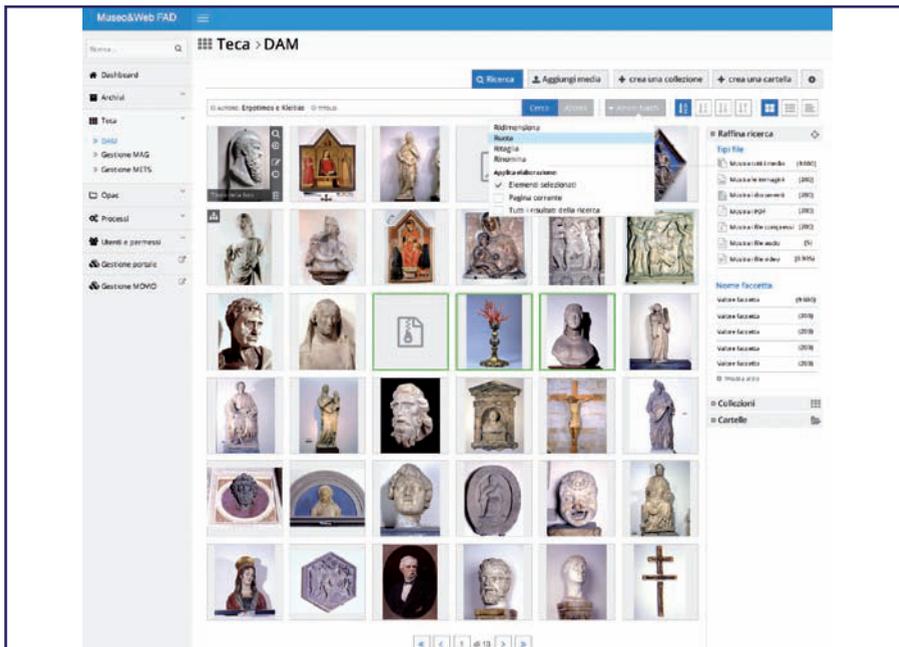


Figura 6: avvio di procedure batch di lavorazione dei file attraverso le procedure integrate nel DAM

Per i metadati strutturali il sistema dispone di interfacce dedicate che consentono all'operatore di stabilire sequenza e nomenclatura dei file componenti l'oggetto della metadateazione. Tali linee procedurali dispongono di un viewer che consente di visualizzare e/o riprodurre i file in modo che l'operatore sia in grado, una volta stabilitane la posizione, di definirne il contenuto attraverso i metadati di nomenclatura (elemento esplicitamente dichiarato nello schema MAG²⁵, che trova corrispondenza anche nel framework METS²⁶).

Una volta definita la struttura fisica, l'operatore passa alla definizione della struttura logica attraverso un'interfaccia grafica che gli consente di aggiungere elementi, anche nidificati, definendone la nomenclatura e la posizione, attraverso la valorizzazione di puntatori agli oggetti digitali la cui sequenza sia stata già definita nella fase preliminare della strutturazione fisica. L'interfaccia dispone a tal fine di un viewer che consente all'operatore di visualizzare i file e definirne i puntamenti.

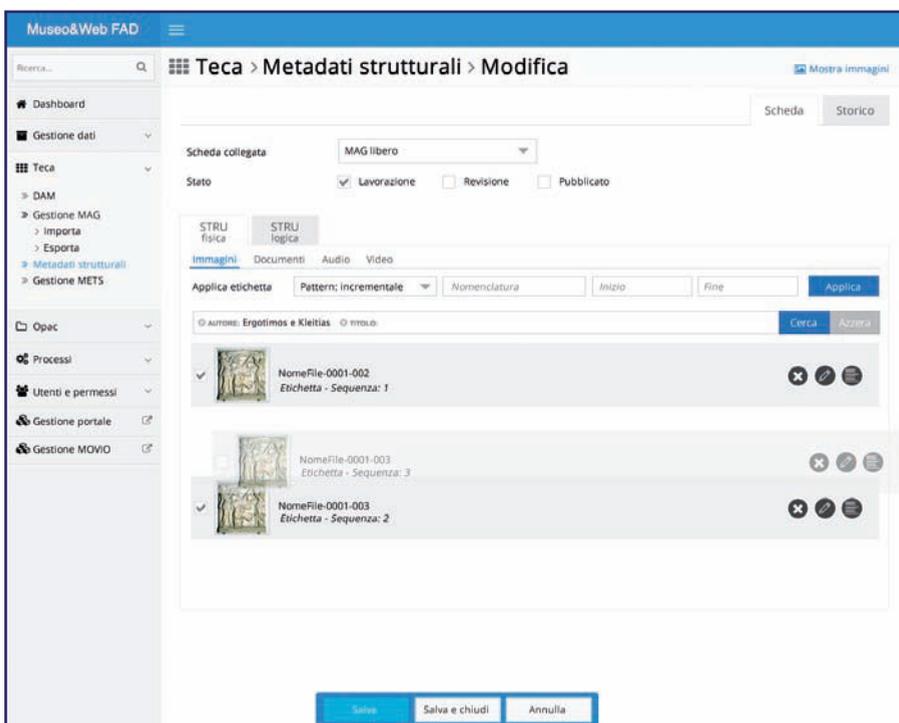


Figura 7: interfaccia di definizione dei metadati strutturali: consente la definizione della struttura fisica di un oggetto complesso e la descrizione della sua tavola dei contenuti (struttura logica)

²⁵ Metadati Amministrativi Gestionali: standard italiano di metadateazione pensato per la gestione per lo più di oggetti digitali generati da campagne di digitalizzazione. Per approfondimenti: <http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati/pagina_267.html>.

²⁶ Metadata Encoding and Transmission Standard: framework standard per la codifica dei metadati descrittivi, amministrativi e strutturali sviluppato dalla Library of Congress. Per approfondimenti: <<http://www.loc.gov/standards/mets/>>.

3.1.1 Schema delle componenti funzionali del sistema sotto il profilo dei servizi digitali disponibili e dei profili di metadati gestiti:

1. DAM (Digital Assets Management o sistema di Digital Library):
 - a. Gestionale di risorse digitali: consente la manipolazione degli oggetti digitali e la generazione dei formati file derivati
 - b. Gestionale di pacchetti (SIP AIP DIP secondo il modello funzionale e informativo OAIS²⁷)
2. Generatore e editor di metadati MAG e METS. Formati gestiti:
 - a. MAG: consente l'esposizione degli oggetti verso l'aggregatore Internet culturale
 - b. METS-SAN: consente l'esposizione dei metadati descrittivi degli allegati digitali verso il sistema informativo del SAN (Sistema archivistico nazionale)
 - c. METS-MDI (MuseiD²⁸)
 - d. METS di Google books: profilo METS generato nell'ambito del progetto Google books

3.2 Gli aggregatori con cui colloquiamo

Lo scopo primario degli schemi di metadati sopra richiamati è quello di rendere ricercabili e usabili gli oggetti digitali quando questi siano esposti fuori dall'ambiente di produzione e gestione originario, di modo che le interfacce applicative degli aggregatori esterni (nel nostro caso Internet culturale e SAN come aggregatori primari) siano in grado di rendere correttamente i pacchetti informativi. I principali schemi di scambio gestiti sono MAG, CAT SAN e METS SAN²⁹: il primo consente il colloquio tramite protocollo OAI-PMH con l'aggregatore Internet culturale verso il quale verranno esposti sia i metadati che gli oggetti digitali; il secondo consente il colloquio con il Sistema archivistico nazionale limitatamente ai metadati descrittivi codificati nella grammatica EAD (Encoded Archival Description) per quanto riguarda la descrizione dei complessi documentale (record di primo livello) ed EAC (Encoded Archival of Contest) per quel che riguarda la descrizione delle entità collegate; gli oggetti digitali e i loro collegamenti ai livelli di descrizione di pertinenza sono descritti attraverso l'uso del profilo METS-SAN che fa uso per il livello descrittivo della codifica EAD mentre per le relazioni si richiama all'ontologia SAN-

²⁷ L'OAIS è un modello di sistema informativo aperto per l'archiviazione di contenuti informativi (Reference Model for an Open Archival Information System), elaborato dal CCSDS (Consultative Committee for Space Data System) e certificato standard ISO 14721:2003. Il documento progettuale, nella forma di raccomandazione CCSDS, è disponibile su: <http://public.ccsds.org/publications/RefModel.asp>.

²⁸ Progetto del Mibac, integrato con CulturalItalia, dedicato ai musei del Paese: http://www.culturalitalia.it/opencms/museid/index_museid.jsp?language=it.

²⁹ Tracciati di scambio che consentono di veicolare rispettivamente metadati relativi alle risorse archivistiche (CAT SAN) e alle unità documentali digitali (METS SAN).

DL³⁰ (SAN Digital Library). Tale sistema, a differenza di Internet culturale, riceverà solo dati descrittivi e non oggetti digitali ai quali l'utente continuerà ad avere accesso attraverso la piattaforma del sistema di Polo (*sistema afferente* nella terminologia del SAN).

La piattaforma del Polo digitale consentirà - grazie ai servizi di metadattazione integrati - un colloquio nativo con i sistemi informativi dei principali aggregatori nazionali. Gli aggregatori di primo livello (Internet culturale per il materiale bibliografico e il SAN per il materiale archivistico) consentiranno di raggiungere - attraverso i canali di interoperabilità garantiti dall'infrastruttura di Cultura Italia - l'aggregatore europeo Europeana.

4. Sistema di gestione dei processi (Workflow Management)

Una componente certamente importante in ambienti di tale complessità è il sistema di modellazione e gestione dei processi di lavoro. Quello presentato sommariamente nel presente lavoro non è un software ma un "sistema di gestione": un insieme di compo-

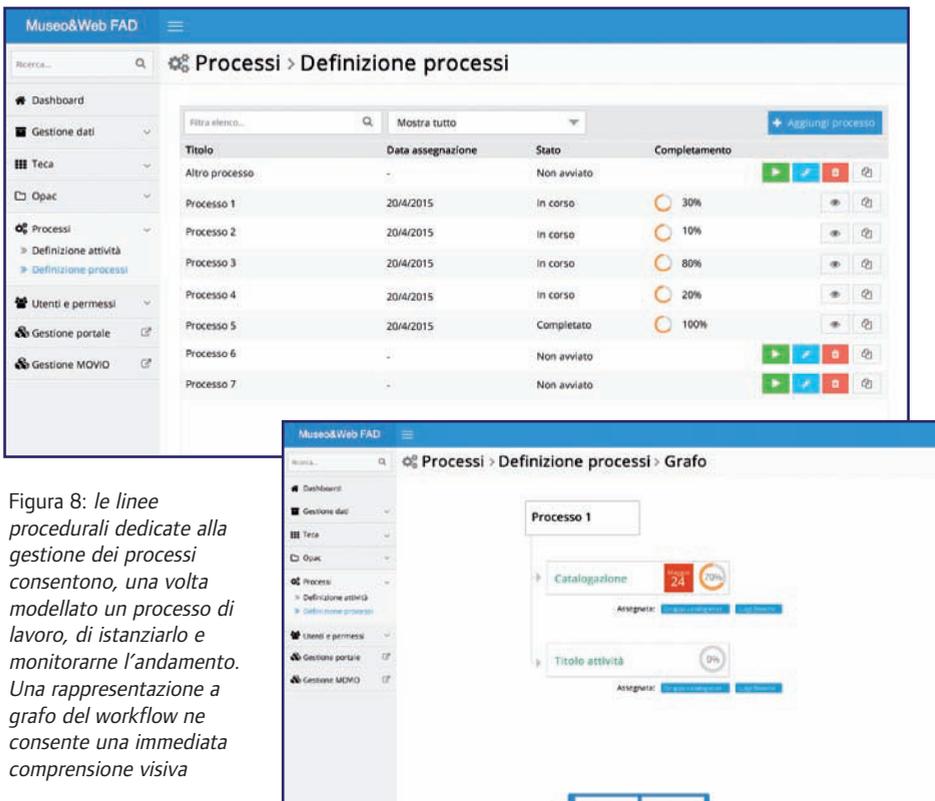


Figura 8: le linee procedurali dedicate alla gestione dei processi consentono, una volta modellato un processo di lavoro, di istanziarlo e monitorarne l'andamento. Una rappresentazione a grafo del workflow ne consente una immediata comprensione visiva

³⁰ Ontologia RDFS che consente di esplicitare, in una sezione dmdSec del METS specializzata, le relazioni che intercorrono tra i documenti digitali descritti e gli altri "oggetti archivistici".

nenti hardware, software, di risorse umane e di regole e procedure da applicare perché il flusso di lavoro finalizzato alla gestione del bene e all'erogazione dei servizi vada a buon fine e sia efficiente. Il componente di gestione dei processi di lavoro consente di modellarne i flussi ottimizzandoli e, quando necessario, ridisegnandoli (reengineering). Il servizio complesso di *document delivery*, per esempio, è gestito attraverso tale componente che consente di monitorare l'intero flusso di lavoro: dalla ricerca effettuata dagli utenti all'acquisizione delle richieste, dalla conseguente movimentazione dei materiali al processo di digitalizzazione in-house dei documenti oggetto di intervento, dalla codifica dei metadati strutturali (fisici e logici) alla disseminazione e alla vendita attraverso le componenti di accesso e restituzione della piattaforma del Polo.

5. Il portale del Polo e il viewer di risorse digitali

Per la realizzazione del portale e dei suoi servizi si useranno le funzionalità di Museo&Web CMS³¹ integrate col sistema di gestione dei dati basato a sua volta su Museo&Web FAD. Grazie al CMS integrato nel sistema di data management, la fruizione dei contenuti provenienti dai cataloghi potrà essere gestita in maniera non convenzionale consentendone una più efficace rappresentazione. Il sito Web collegato al sistema di catalogazione/descrizione del patrimonio, attingendo quindi direttamente ai repositories documentali, permetterà una rappresentazione dei dati descrittivi di catalogo e degli oggetti digitali collegati sotto forma di percorsi tematici, collezioni di opere e itinerari di visita collegati a interfacce geografiche. Il passaggio dal dato catalogato alla sua resa all'interno di percorsi di presentazione graficamente accattivanti e orientati alla didattica è certamente un punto di forza reso possibile dall'integrazione tra i due sistemi.

Un componente fondamentale in un sistema di digital library è certamente il visualizzatore di risorse digitali. Pensato come plug-in integrato nella piattaforma modulare – ma usabile anche separatamente perché basato su profili di metadati standard – si occupa del recupero e della presentazione degli oggetti digitali e dei metadati (descrittivi e strutturali) associati tramite DAM o contenuti nell'oggetto fisico lato server. A tale componente, oltreché all'OPAC, è affidato inoltre il compito di rendere disponibili agli utenti gli strumenti necessari a effettuare le richieste di servizi sul patrimonio: quello già digitalizzato, attraverso il download integrale o selettivo dei media, a pagamento o free che siano; quello non ancora dematerializzato attraverso le richieste di digitalizzazione – effettuate direttamente da OPAC – grazie all'apposita modulistica (RIA³²).

L'erogazione di tali servizi ha richiesto, in fase di progettazione, lo studio di un'opportuna integrazione funzionale tra le componenti applicative coinvolte nel

³¹ Content Management System (CMS) pensato per facilitare la creazione e l'aggiornamento online di siti Web destinati alla valorizzazione del patrimonio e delle istituzioni culturali. Per approfondimenti: <<http://www.minervaeurope.org/structure/workinggroups/userneeds/prototipo/cms.html>>.

³² Rich Internet Application: formulari dinamici online che consentono agli utenti di effettuare le richieste di servizi passando agli operatori di back-office tutti i riferimenti necessari ad avviare i

processo di acquisizione delle richieste: l'OPAC con il viewer e la modulistica RIA; questi ultimi con il server di e-commerce incluso nella fornitura.

L'interfaccia di accesso e richiesta così configurata consente di acquisire in modo dinamico i riferimenti e i parametri necessari ad avviare il processo di *document delivery*, modellato e gestito attraverso il sistema di workflow management (vedi paragrafo 4).

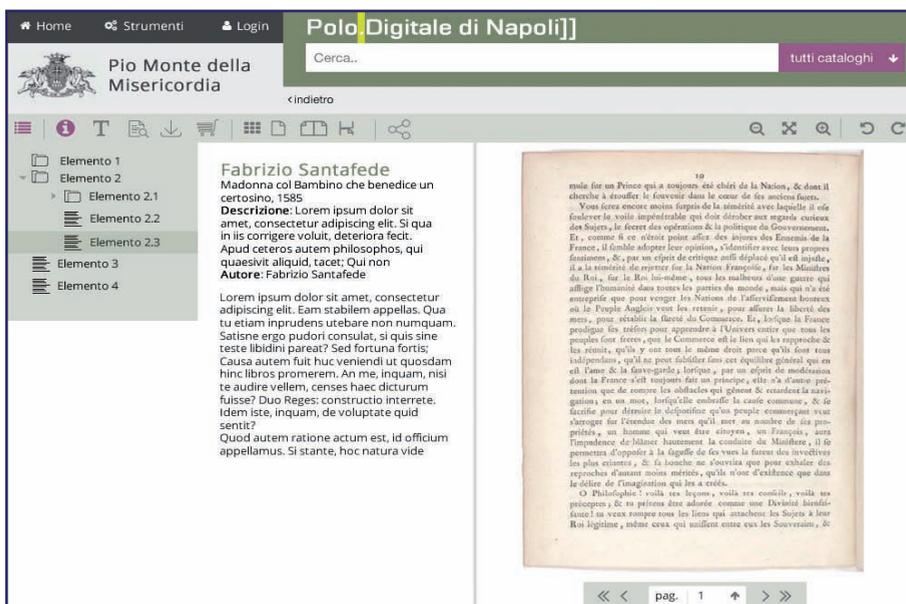


Figura 9: il visualizzatore integrato rappresenta la componente cui è demandato non solo il compito di rendere visibili le risorse digitali, ma anche di consentire all'utenza di avviare richieste di servizi

6. Sistema di gestione semantica e LOD (Linked Open Data)

Tra gli obiettivi più ambiziosi del progetto vi è lo sviluppo di un sistema di gestione semantica in grado di accompagnare il Polo nel complesso processo della produzione e del mantenimento dei dati aperti linkati LOD³³. Lo scopo del sistema di gestione semantica del Polo digitale è la produzione dinamica di dataset tramite gli appositi linguaggi costituiti dal modello dei dati RDF³⁴ (per la strutturazione sintattica dell'informazione) e le ontologie di dominio scelte nell'ambito del progetto (come approfondito nel prosieguo a proposito dell'OPAC di Polo, il modello FRBRoo³⁵ per i domini bibliografico e museale e il thesaurus di concetti archivistici

loro processi di lavoro, escludendo da tale flusso l'uso della posta elettronica e consentendo in tal modo al personale di risparmiare tempo, eliminando compiti ridondanti come verificare la correttezza delle informazioni riportate.

³³ <http://www.w3.org/wiki/SweolG/TaskForces/CommunityProjects/LinkingOpenData>.

³⁴ <http://www.w3.org/RDF/>.

³⁵ Functional Requirements for Bibliographic Records - object oriented: <http://www.cidoc-crm.org/frbr_intro.html>.

OAD³⁶ per il dominio archivistico). I dataset generati – a partire dalle fonti dati interne o esterne al sistema (come nel caso dei dati SBN recuperati in cooperazione applicativa) e modellati sulla base delle ontologie di dominio – saranno collegati a dataset esterni come Dbpedia³⁷ (la versione opendata di Wikipedia, ad esempio per le schede biografiche di autori che potranno essere in tal modo dinamicamente richiamate), GeoNames³⁸ (per la georeferenziazione dei luoghi, riferibili ai luoghi di edizione di una pubblicazione, ma anche ai luoghi citati nei documenti digitalizzati) e VIAF³⁹ (Virtual International Authority File per le intestazioni di authority).

6.1 Arricchimento semantico e NER (Named Entity Recognition)

L'arricchimento semantico sarà utilizzato per collegare alcuni metadati contenuti nelle schede di catalogazione con risorse esterne (le ontologie dei principali dataset sopra richiamati) che potranno consentire approfondimenti o una navigazione più efficace. Questi collegamenti saranno aggiornati in tempo reale e consentiranno di avere una informazione sempre allineata alle fonti esterne. In sostanza il sistema di *semantic enrichment* analizzerà il contenuto del metadato con tecniche di NLP (Natural Language Processing), lo normalizzerà e creerà un metadato RDF che sarà esportato come tripla utilizzando una proprietà specifica come *sameAs*, il collegamento alla risorsa esterna sarà risolto in tempo reale attraverso l'URI della risorsa stessa. Tale tecnica consentirà di alimentare con un processo automatico il linking esterno con i principali dataset sopra richiamati.

Le tecniche di *semantic enrichment* saranno applicate anche al testo trascritto, o all'OCR, con lo scopo di estrarre dall'informazione non strutturata quei metadati che potranno essere usati per descrivere la risorsa in modo semiautomatico.

7. OPAC (Online Public Access Catalogue) e base di conoscenza

Pur essendo tali domini formalmente separati e gestiti sulla base di normative specifiche (SBN: per il materiale bibliografico; SAN - ISAD: per la documentazione d'archivio; normative ICCD per la descrizione del patrimonio artistico), sono diverse le relazioni e, in parte, le sovrapposizioni tra gli oggetti che costituiscono tali domini: gli archivi, residuo e traccia documentale dell'attività istituzionale di un soggetto produttore, possono testimoniare della formazione delle raccolte d'arte; le opere d'arte sono l'oggetto di studi che costituiscono intere bibliografie. Tale modello ontologico costituisce la base di conoscenza imprescindibile per costruire un punto di accesso unificato, il cui modello è rappresentato dalle ontologie di dominio FRBRoo e OAD (di cui tratteremo brevemente più avanti), impiegate nel progetto.

³⁶ Ontology Archival Description: <<http://labs.regesta.com/progettoReload/wp-content/uploads/2013/08/oadNew.html>>.

³⁷ <http://it.dbpedia.org/>.

³⁸ <http://www.geonames.org/>.

³⁹ <https://viaf.org/>.

7.1 La comunità degli utenti (interni ed esterni):

Per chi impiantare un sistema di questo tipo? Due sono le classi di utenti:

1. gli utenti esterni: gli studiosi, gli studenti, i lettori e i cultori che orbitano normalmente intorno a tale tipo di istituti culturali. Questa comunità di riferimenti è caratterizzata da esigenze informative molto differenziate:
 - a. le esigenze specialistiche dello studioso sono esaudite dai cataloghi di dominio: SBN per la ricerca bibliografica finalizzata all'identificazione e localizzazione dei documenti a stampa (tipica funzione del catalogo); i tracciati (di livello inventariale, pre-catalogo e catalogo) definiti dalle normative ICCD adottate per la descrizione del patrimonio, che consentono di accedere a una descrizione più dettagliata, più analitica del bene (finalizzata alla sua conservazione); e, infine, la descrizione dei complessi archivistici la cui strutturazione gerarchica, analiticità e ricchezza informativa (il modello dati è basato sullo standard internazionale ISAD) garantisce la contestualizzazione dell'informazione veicolata dalla documentazione d'archivio.
 - b. le esigenze informative dell'utente generico sono esaudite dal sistema di meta-indice: identificare risorse informative di qualsiasi natura correlate in qualsiasi modo a un oggetto, un evento, una persona, una famiglia, un ente. È a questo tipo di ricerche integrate che si rivolge l'interfaccia di ricerca *cross-domain* dell'OPAC.
2. Le esigenze degli utenti interni o operatori delle istituzioni coinvolte: bibliotecari, archivisti, addetti di sala, curatori, dal lavoro dei quali dipende la gestione dei servizi all'utenza. L'introduzione della piattaforma di Polo consente l'ottimizzazione e, dove necessario, la reingegnerizzazione dei processi di lavoro in essere. Il *reengineering* dei processi è infatti uno degli obiettivi primari del progetto (vedi paragrafo 4. Sistema di gestione dei processi).

7.2 Un'interfaccia di ricerca innovativa (FRBR) e l'orientamento ai servizi

Nello scenario attuale il lettore riceve, in risposta a un'interrogazione, un elenco di record bibliografici che descrivono pubblicazioni sulle quali bisogna poi operare una selezione, talvolta complessa, per individuare le informazioni di interesse: gli oggetti che un sistema tradizionale restituisce sono tutti della stessa natura, sono edizioni pubblicate di date opere.

Il nuovo modello di OPAC intende rispondere in modo più puntuale alla richiesta di ciascun utente; la risposta mirata è un vantaggio che deriva dall'adozione della struttura dei dati organizzata sulla base del modello FRBR (Functional Requirements for Bibliographic Records), che distingue le diverse entità – attraverso le quali è organizzabile concettualmente il dominio della conoscenza bibliografica e museografica – e che corrispondono all'approccio della ricerca: un utente, infatti, può essere interessato ad avere informazioni su un'opera (il contenuto artistico o intellettuale identificato da una concatenazione di titolo e autore); può essere interessato a una particolare

espressione dell'opera (cioè a un determinato testo, es. una certa traduzione o una determinata edizione dell'opera); oppure può ricercare una particolare manifestazione bibliografica (l'oggetto dei cataloghi tradizionali) della quale conosce dati precisi, come il titolo, il nome dell'autore, la data di pubblicazione ecc. Questa puntualità nella risposta è possibile perché lo standard guida cui ci riferiamo prevede la produzione granulare di metadati abbinati alle entità del modello: opera, espressione, manifestazione, item; oggetto, concetto, evento, persona, famiglia, ente.

7.2.1 Schema delle componenti funzionali del sistema di OPAC

Il sistema prevede l'integrazione dei seguenti componenti software:

1. *Sistema di indicizzazione delle schede secondo i seguenti modelli dei dati:*
 - a. EAD-EAC (dominio archivistico)
 - b. SBNMarc (dominio bibliografico affidato per la fase di record making all'ambiente SBN)
 - c. ICCD (F, D, S, OA)
 - d. MODS
 - e. TEI-MS (Manus compliant)
2. *Sistema di costruzione dell'indice di metaopac*
 Questo componente è l'interfaccia di amministrazione in Museo&Web FAD delle modalità di fruizione dell'OPAC per ogni tipologia di scheda e per il profilo comune. È possibile definire sia i campi di ricerca per la modalità "google like" che quelli per la ricerca avanzata.
3. *Modulo di ricerca a faccette*
 I risultati delle ricerche, sia quelle semplici che avanzate, saranno presentate corredate di faccette che consentiranno di navigare i risultati apponendo dinamicamente gli opportuni filtri definiti sulla base dei modelli gestiti.
4. *Visualizzazione sintetica e dettagliata*
 Per ogni tipologia di scheda si prevede una visualizzazione dei dati in forma sintetica che consentirà di presentare i dati più rilevanti insieme all'anteprima del documento (se presenti oggetti digitali collegati). Tutti i dati associati all'oggetto saranno visibili poi in pagine di dettaglio che ne offriranno una vista analitica organizzata per sezioni.
5. *Modulo FRBR*
 Tale componente costituisce il GUI layer (livello di presentazione dei dati) che consentirà agli utenti una navigazione del dataset RDF costruito sulla base del modello ontologico FRBRoo a partire dalla fonte dati UNIMARC di SBN. Tali interfacce consentiranno una navigazione a percorsi pre-definiti tra le entità del modello.

Le interfacce di amministrazione saranno realizzate in javascript con la tecnologia Angularjs, il motore di ricerca sarà sviluppato in SOLR e i servizi da esporre per la gestione delle interfacce di consultazione in PHP.

8. Conclusioni e prospettive

Obiettivo immediato del progetto è la creazione di un ambiente gestionale integrato che supporti le attività di valorizzazione⁴⁰ del patrimonio culturale degli istituti partner. Tali attività sono certo rese possibile dall'implementazione della piattaforma software; ma in primo luogo richiedono la formazione del personale sui complessi processi di lavoro finalizzati alla digitalizzazione e divulgazione di beni culturali documentali e museali.

Tra ICCU e **Polo digitale degli istituti culturali di Napoli** – come accennato in apertura del presente intervento – è attiva una collaborazione che va avanti dal febbraio 2015. Tale cooperazione, dall'intento originario limitato allo sviluppo in formato LOD dei dati bibliografici localizzati in SBN dagli istituti partner del Polo digitale, si è estesa al rilascio di una piattaforma software, gratuitamente disponibile, che consentirà alle istituzioni della cultura di sostenere l'intero ciclo di produzione e gestione del digitale, finalizzato all'erogazione di servizi destinati alla comunità dei propri utenti.

Tale obiettivo è stato formalizzato nella **Convenzione**⁴¹ firmata nel gennaio 2016, nella quale sono stati definiti gli impegni reciproci. Da un lato la collaborazione tecnico-scientifica dell'ICCU che si esprime sia nel lavoro di estensione e raffinamento delle mappatura tra il modello dati di SBN e l'ontologia di dominio (FRBR) adottata per lo sviluppo in LOD dei dataset bibliografici, sia nel garantire le evolutive applicative dell'ambiente SBN (protocolli e servizi software) utili a favorire il colloquio in cooperazione applicativa, in particolare tra SBNWeb e Museo&Web FAD. Dall'altro l'impegno del Polo digitale che si manifesta attraverso il coinvolgimento dell'ICCU nella fase finale dello sviluppo e dell'ottimizzazione della piattaforma e nella garanzia della piena disponibilità della stessa in modo che l'ICCU possa sia riusarla a livello centrale che promuoverne la diffusione verso tutte quelle istituzioni interessate a implementare un proprio sistema di biblioteca digitale.

⁴⁰ Da intendersi qui non come monetizzazione del bene – sebbene sia disponibile anche un sistema di e-commerce – ma come accesso e conoscenza che presuppone la descrizione, lo studio e, non ultimo, la tutela del bene.

⁴¹ http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2016/Convenzione_ICCU_polo_digitale_istituti_cilturali_Napoli.pdf

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

La banca dati del manifesto politico

Siriana Suprani

Fondazione Gramsci Emilia-Romagna

La banca dati www.manifestipolitici.it nasce da un progetto della Fondazione Gramsci Emilia-Romagna di oltre dieci anni fa. Propone alla consultazione online una raccolta di circa 13.000 (in progressivo aumento) manifesti politici e sociali dal Novecento ad oggi. Passano sullo schermo del computer le immagini dei grandi fogli di carta che tappezzavano i muri e che, con la costituzione dei partiti di massa nel secondo dopoguerra, sono stati strumento di propaganda e di comunicazione dei partiti e dei movimenti politici.

Ho usato indifferentemente i due concetti, quello di propaganda e di comunicazione. Gli storici si sono posti invece la domanda se il loro significato sia analogo e cioè se il termine propaganda debba avere solo un'accezione negativa e alluda solo a «pratiche politiche di tipo manipolativo e prevaricatorio. Ciò ha determinato un abbandono del termine a favore di altri quali pubblicità, marketing, relazioni pubbliche o anche più semplicemente comunicazione politica»¹. C'è del vero anche se io penso che se ci riferiamo ad un ambito politico e sociale dove vigono dinamiche democratiche e pertanto dove più voci competono e si confrontano, con libertà di decodificazione dei meccanismi, il termine propaganda sia solo storicamente datato e si possa considerare equivalente nella sostanza a quello di comunicazione politica. In ogni caso rinvio per approfondimenti al volume *Propagande contro* curato da Andrea Baravelli per i caratteri della casa editrice Carocci.

Il manifesto è stato uno dei media più utilizzati soprattutto prima dell'affermarsi della televisione. Molti documenti, come ad esempio *Il quaderno dell'attivista* periodico del PCI dei primi anni del secondo dopoguerra, testimoniano come al militante di partito si insegnasse come realizzare un manifesto, dal linguaggio all'immagine, come gli venissero indicati i luoghi più idonei per l'affissione e come scegliere i destinatari.

Il manifesto contemporaneo, soprattutto quello nel pieno della sua colorata espressione che va dal dopoguerra agli anni ottanta, è solo in parte l'erede dei bandi e dei fogli volanti, che presero a circolare fin dai tempi dell'avvento della stampa per diffondere informazioni appelli e inviti, e si affermarono in Italia sotto l'influenza della rivoluzione francese e della occupazione napoleonica².

¹ Andrea Baravelli (a cura di), *Propagande contro. Modelli di comunicazione politica nel XX secolo*, Roma, Carocci, 2005.

² *Fogli volanti toscani: catalogo delle pubblicazioni della Biblioteca di storia moderna e contemporanea di Roma 1814-1849* a cura di Sara Mori, Milano, Franco Angeli, 2008.