

Antonella Negri - Jacopo Russo

Analisi ed elaborazione dei dati catalografici per l'allineamento agli standard ICCD

Con l'entrata in esercizio del SIGECweb l'attività di catalogazione può essere svolta direttamente *on line* semplificando significativamente il flusso di lavoro. Esiste tuttavia un importante patrimonio di dati prodotti un tempo in forma cartacea e poi in formato digitale, con differenti sistemi informativi.

Gli standard catalografici predisposti da ICCD non vincolano il catalogatore all'uso di uno strumento specifico. Per permettere lo scambio di informazioni tra i diversi sistemi informativi fu messo a punto uno standard di trasferimento che permettesse di importare ed esportare attraverso file di testo, il contenuto di complesse banche dati di tipo relazionale. La maggior parte del materiale consegnato a ICCD in passato ed ancora oggi è nel formato testuale denominato "ICCD92".

L'adozione di questo standard ha permesso di lavorare con grande flessibilità. Per contro una non corretta applicazione del formalismo previsto e un'interpretazione erronea degli standard catalografici portano a serie difficoltà in fase di validazione ed acquisizione dei dati nel Catalogo Nazionale.

Molto spesso le schede, consegnate in un passato più o meno recente, non possono essere caricate perché contengono errori che ne impediscono l'acquisizione. Questo problema ha creato un ostacolo al flusso previsto per la catalogazione. I dati provengono da fonti estremamente eterogenee e dunque la casistica delle anomalie che possono contenere è molto variegata. Le difficoltà incontrate nella validazione di materiale proveniente dall'intero territorio nazionale hanno portato nel tempo ad accumulare un importante pregresso. Per queste ragioni ci si è recentemente dotati di strumenti di analisi utili a individuare in maniera puntuale i problemi presenti sui dati e a decidere quali strategie adottare per superarli ed acquisire le schede.

Quando ICCD riceve e protocolla un plico proveniente da un ente schedatore, i dati vengono presi in carico ed inizia un iter per la loro revisione. Questa attività è stata resa sistematica ed è informatizzata.

Quando una consegna è presa in carico si compila una scheda con i dati dell'ente, le date, i protocolli ed i referenti che hanno curato il lavoro. Il sistema crea automaticamente due cartelle associate alla consegna: una per gli originali e l'altra per la revisione.

Il contenuto di ciascun supporto consegnato viene copiato in apposite cartelle; una procedura automatica crea un archivio compresso per ciascun supporto, quindi sposta i dati nella cartella di revisione. I file originali (pacchetti di schede, authority files, documentazione allegata, ed ogni

Antonella Negri - Jacopo Russo

Analisi ed elaborazione dei dati catalografici per l'allineamento agli standard ICCD

altro file trasmesso) vengono indicizzati “fotografando” il contenuto di ciascun supporto ed archiviati; da questo momento in poi non saranno più modificati e restano a disposizione per consultazione.

La revisione dei dati catalografici prevede una serie di test eseguiti a cascata, l'uno sull'output del precedente. L'analisi vera e propria è preceduta da una fase preliminare in cui si verifica l'integrità dei caratteri di ritorno a capo e la correttezza dell'intestazione. Se un file è privo di intestazione valida questa viene rigenerata, laddove possibile, sulla base dei dati riportati nelle schede. La procedura di test riconosce automaticamente i file contenenti dati catalografici e, se questi hanno un'intestazione valida, cerca nel proprio database una normativa compatibile da associare a ciascun file per verificare le schede. L'applicazione è in grado di processare un intero albero di cartelle e sottocartelle con molti file di diverse normative attraverso una procedura batch.

Verifica della struttura

Il primo test eseguito è quello sulla struttura del file. Si tratta del controllo più delicato e propedeutico ai test successivi. L'implementazione di questa procedura ha richiesto il maggior sforzo perché il file, all'inizio dell'analisi potrebbe contenere virtualmente qualsiasi tipo di errore. La struttura potrebbe essere danneggiata più o meno gravemente, al limite il file potrebbe non contenere alcuna struttura catalografica riconoscibile. Il test procede analizzando il file riga per riga, verificando la correttezza dei formalismi e le relazioni gerarchiche tra i campi riportati in righe sequenziali. Al termine del test eventuali errori vengono corretti in maniera da poter derivare dal file una struttura compatibile con lo standard dichiarato nell'intestazione. In questa fase vengono intercettati e corretti i seguenti tipi di errore:

- errori nelle relazioni gerarchiche;
- errori di sequenza o ripetizione in sottocampi di strutture ripetitive;
- campi estranei alla normativa dichiarata nell'intestazione;
- schede estranee alla normativa dichiarata all'intestazione;
- errori nell'intestazione delle righe;
- righe vuote;
- righe costituite di soli spazi;
- righe più lunghe di 80 caratteri;
- campi semplici non valorizzati;
- campi strutturati o paragrafi valorizzati.

Antonella Negri - Jacopo Russo

Analisi ed elaborazione dei dati catalografici per l'allineamento agli standard ICCD

La correzione automatica di alcune di queste tipologie di errore comporta una perdita di contenuti informativi. Quando questi errori vengono rilevati si procede in ogni caso per verificare, attraverso i successivi test, se le schede contengono ulteriori tipologie di errore.

Quando un file non contiene errori nella struttura è possibile sottoporlo ai successivi test.

Sequenze: Il test sulle sequenze intercetta e corregge automaticamente errori nell'ordine di campi semplici ma anche si intere strutture o paragrafi all'interno della scheda. La correzione di questi errori non comporta una perdita di contenuti informativi.

Duplicati: Il test sui duplicati verifica che la ripetitività sia applicata interpretando correttamente la normativa. Se un campo semplice, una struttura o un paragrafo presenta più istanze, laddove la normativa non prevede la ripetitività, l'applicazione segnala un errore. La correzione automatica conserva la prima istanza eliminando le successive; questa correzione comporta una perdita di dati.

Obblighi: Il test controlla, a partire dalla normativa, se nella scheda sono compilate le obbligazioni assolute e contestuali. Se un campo non è presente con la correzione automatica questo viene inserito, valorizzandolo con la stringa "NR". Se una struttura o al limite un intero paragrafo non sono presenti, la procedura opera ricorsivamente compilando l'intero ramo con tutti i campi obbligatori, valorizzandoli con "NR". Nel caso delle obbligazioni alternative (implicite o esplicite) viene inserito il primo dei campi tra i quali è prevista la scelta. Questo tipo di correzione non porta a una perdita di informazioni ma conduce a una valutazione erronea in merito alla completezza della scheda. Il valore inserito inoltre può risultare incompatibile con eventuali vocabolari. La valorizzazione dei campi con queste stringhe è funzionale all'importazione delle schede in un'attività di revisione per una successiva modifica manuale.

Lunghezze: Questo test verifica che le lunghezze dei valori inseriti nei campi siano minori o uguali a quanto previsto dalla normativa. Stringhe eccedenti vengono tagliate. Prima di eseguire questo test si verifica che le stringhe non contengano degli spazi superflui a destra, che porterebbero a dei falsi positivi nel test sulle lunghezze. La correzione automatica di questo errore porta ad una perdita di contenuto informativo.

Al termine del lavoro si compila un rapporto esaustivo sugli eventuali errori rilevati. Se gli errori sono meramente formali e la loro correzione non comporta alterazioni del contenuto, il materiale può essere importato direttamente. Se invece la correzione necessaria per rendere il materiale conforme è tale da alterare o depauperare il contenuto informativo delle schede, il rapporto viene

Antonella Negri - Jacopo Russo

Analisi ed elaborazione dei dati catalografici per l'allineamento agli standard ICCD

trasmesso all'ente che ha consegnato i dati per una richiesta di correzioni o integrazioni o per un parere sulle correzioni proposte per caricare il materiale, predisponendolo ad una successiva revisione manuale. Eventuale nuovo materiale consegnato viene preso in carico ed analizzato perfezionando il rapporto.

La verifica dei dati catalografici è completata da un controllo sulla composizione e consistenza formale del pacchetto. Tutti i contenuti del pacchetto (schede, authority file, documentazione allegata ed altri file a corredo) vengono registrati. Si verifica che il file di collegamento con gli allegati non presenti anomalie, generando eventualmente una reportistica di dettaglio con gli errori rilevati. Se sono presenti disallineamenti tra gli allegati dichiarati nel file immftan.txt e quelli effettivamente presenti all'interno del pacchetto, viene generata una tabella di confronto con gli eventuali allegati dichiarati ma mancanti ed i file inclusi nel pacchetto ma privi di riferimento alle schede.

Le fattispecie di errori presenti nei pacchetti sono, come è facile intuire, molto varie. Talvolta un pacchetto può contenere pochi errori la cui correzione può essere effettuata in automatico, in altri casi gli errori possono essere tali da richiedere una revisione manuale. Lo scopo della revisione è quello di fare emergere tutti i problemi presenti, per questa ragione i test vengono sempre condotti tutti, eventualmente anche introducendo delle forzature attraverso la correzione automatica.

Il risultato del procedimento di revisione è un pacchetto importabile da un punto di vista meramente formale. Un'attenta valutazione basata sull'esito dei test porterà alla scelta, di concerto con l'ente che ha prodotto i dati, tra diverse alternative quali:

- importare i dati accettando le correzioni automatiche;
- importare i dati predisponendoli per un'attività di revisione;
- revisionare i dati manualmente prima di procedere all'importazione.

La revisione costituisce un importante *feedback* per gli enti impegnati nella catalogazione. Laddove vengano riscontrati errori sistematici, la revisione dei dati e la reportistica che viene generata, possono essere un valido aiuto per gli sviluppatori che curano la manutenzione di sistemi informativi diversi da SIGEC. Spesso un disallineamento dagli standard catalografici può generarsi a valle, per un errore nelle procedure di export che producono i pacchetti in formato ICCD92. In tutti questi casi l'analisi sistematica dei dati è utile non solo per il recupero di quanto prodotto in passato, ma anche per ottimizzare il lavoro futuro evitando uno spreco di risorse.