

PROGETTO SRI

Gruppo di lavoro 2

Definizione delle mappature per la costruzione del Cluster di authority Conclusioni del lavoro del gruppo

PREMESSA	1
STRUTTURA E MAPPATURA DEI DATI	2
EDIT16	2
MANUS ONLINE	2
CONTENUTO E CONFRONTO DEI DATI	3
EDIT16	3
MANUS ONLINE	3
CONCLUSIONI	4

Premessa

Il gruppo 2, composto da Massimo Gentili-Tedeschi (coordinatore), Carla Baccini, Flavia Bruni, Giuseppina Leolini, Paola Lolli, Lucia Negrini, Elena Ravelli, Margherita Aste, Valentina Atturo, Carla Scognamiglio, ha avuto il compito di definire le mappature per la costruzione del cluster di authority. I lavori sono iniziati il 23 ottobre 2017 e si sono conclusi il 13 novembre.

Il gruppo ha organizzato il lavoro suddividendosi in due sottogruppi con i seguenti compiti:

1. mappatura dei tag e degli elementi dei dati di autorità (nomi e luoghi) di SBN, Manus-On-Line e Edit16, tenendo conto non solo del protocollo SBNMarc ma anche del formato UNIMARC standard (Atturo, Baccini, Gentili-Tedeschi, Lolli, Negrini, Ravelli);
2. controllo degli autori personali e collettivi identificati di MOL sulla base di un campione di circa 200 nomi in forma accettata (nomi A, B, C, D, E, R, G) per valutare la possibilità di un confronto e dell'integrazione degli archivi (Aste, Bruni, Gentili-Tedeschi, Leolini, Negrini, Scognamiglio).

In MOL sono presenti circa 151.000 nomi identificati e 52.000 non identificati; includendo forme varianti e nomi presenti sui manoscritti, 267.000 nomi.

In Edit16 sono presenti circa 13.000 nomi, oltre a 2.600 nomi di tipografi, gestiti in un archivio separato.

Struttura e mappatura dei dati

Edit16

Il formato dei dati dei nomi di Edit16 è sostanzialmente uguale a quello di SBN, con le seguenti differenze:

- l'identificativo del record è CNCA seguito da 6 cifre, ma quando vi è corrispondenza con un VID di SBN la base dati gestionale riporta anche un campo con l'identificativo SBN.
- come identificativo del nome si usa l'ISADN (015) invece dell'ISNI (010); l'ISNI è in corso di implementazione.
- in Edit16 manca il campo Lingua (101).
- in Edit16 il campo Regole di catalogazione (152) è RICA, non REICAT.
- in Edit16 le date di nascita e di morte (nomi personali) sono correttamente gestite nel sottocampo 200 \$f, mentre in SBN sono in una nota generale (300 \$a).
- in Edit16 e in SBN le date relative all'attività degli enti sono erroneamente gestite nel campo 210 \$f, non nel campo 640 \$f,\$i.
- in Edit16 il nome presente sulle edizioni è gestito in un campo locale 309, non presente in SBN
- in Edit16 le note biografiche sono correttamente gestite nel campo 340, mentre in SBN si usa erroneamente il campo 300.
- in Edit16 le informazioni aggiuntive sul tipografo sono gestite nel campo 349, corrispondente al campo standard UNIMARC 341 non presente in SBN.
- in Edit16 il luogo di pubblicazione normalizzato è gestito come in SBN con l'etichetta 260 \$a,\$d (nazione e città), ma sarebbe auspicabile implementare anche i sottocampi per entità territoriali intermedie.
- l'etichetta 810 per la fonte bibliografica consultata con esito positivo è gestita in Edit16 come in SBN, ma in entrambi non è chiaro se in SBN le informazioni ricavate dalla fonte siano correttamente inserite nel sottocampo \$b.
- in Edit16 le fonti consultate con esito negativo non vengono gestite nel campo 815 ma nella nota del catalogatore (830).
- in Edit16 le fonti consultate disponibili in formato elettronico sono linkate con un campo locale 899, non presente in SBN e senza un equivalente in UNIMARC.

Manus OnLine

Ancorché espressi con etichette diverse da SBNMarc e da UNIMARC, i dati registrati nella scheda di autorità (SchedaNomiMOL) di MOL sono compatibili con quelli registrati in SBN:

- il Tipo nome SBN di MOL equivale al Tipo nome di SBN (UNIMARC 200 indicatore 2, nome in forma diretta o inversa).
- l'identificativo del record (CNMN) è numerico a 10 cifre.
- i nomi sono suddivisi tra forma accettata e identificata e forma accettata non identificata, che si potrebbero allineare a Livello 97 e 05 di SBN.
- le forme di rinvio sono gestite in modo analogo a SBN.
- l'identificatore del nome è lo stesso codice ISNI.
- il Nome è gestito in modo sostanzialmente uguale.
- la Qualifica è equivalente alla qualificazione in SBN.

A_8) SRI_DLG_CTecnico_relazione_conclusiva_dati_di_autorita

- la Datazione corrisponde alle Datazioni in SBN.
- in MOL è presente un campo Importazione Qualifica + Datazione, utilizzato nel 2007 per riversare i dati da Manus 3 su PC a ManusOnLine, ma se ne prevede l'eliminazione.
- in MOL esiste un campo Lingua prevalentemente utilizzato, corrispondente al campo Lingua (101) di SBN, ma il dato testuale non è controllato.
- in MOL esiste un campo Note al nome, Fonti o note bibliografiche in cui i dati equivalenti a Nota informativa (340), Nota biografica (810) e Note del catalogatore (830) sono indistinti; in SBN la Nota informativa (300), le Fonti (810) e la Nota del catalogatore (830) sono distinte.
- in MOL manca un codice Paese (102).
- in MOL sono registrati Nomi presenti sul manoscritto, collegati al formato bibliografico.

Contenuto e confronto dei dati

Edit16

Non si è ritenuto utile entrare in un esame a campione dei nomi presenti nei due archivi, ma si sono rilevate le seguenti problematiche perché un allineamento possa risultare attendibile:

- le procedure batch di confronto tra record di Indice e record di Edit16 individueranno soprattutto per l'archivio editori una grande quantità di nomi simili che occorrerà esaminare puntualmente e manualmente.
- per i nomi simili e per quelli non qualificati il confronto dovrà essere esteso alle registrazioni bibliografiche collegate al fine di un'identificazione certa.

ManusOnLine

Un esame a campione fatto su un set di circa 1400 nomi identificati di MOL confrontati con SBN ha evidenziato le seguenti problematiche:

- molti nomi presenti in MOL non risultano presenti nell'Indice SBN;
- le forme accettate a volte si presentano in forma diversa (Giotto : di#Bondone - MOL; Giotto <1267?-1337> - SBN; Thomas : More – MOL; More, Thomas <1478-1535> - SBN);
- le qualificazioni sono quasi sempre presenti ma la loro sintassi è diversa rispetto a quella di SBN (si ricorda che nelle procedure di confronto le qualificazioni contano);
- il codice del tipo nome a volte è utilizzato in modo errato (Farfa e nomi di famiglie registrati con tipo nome 'A'; Cyrillus <santo>, Paulus : de#Sancta Maria <m. 1435>, Gualfreducci, Bandino <1565-1627> con tipo nome 'B') perché il software precedentemente non effettuava alcun controllo;
 - per i nomi 'C' ci sono meno problemi, ma:
 - spesso la forma accettata è quella variante di SBN;
 - le qualificazioni sono quasi sempre diverse da quelle SBN (in MOL si riportano sempre le qualificazioni cronologiche, ad es. Alighieri, Dante in SBN e Alighieri, Dante <1265-1321>) in MOL;
 - le date in SBN a volte sono visualizzate nella nota informativa, quando inserite con applicativi diversi da Sbnweb e da Interfaccia Diretta;
- tra i nomi di tipo 'E' sono presenti ad es.: *Mostra del *Tintoretto <1937 ; Venezia>, *Doria <famiglia>, The *Times *literary *supplement;

A_8) SRI_DLG_CTecnico_relazione_conclusiva_dati_di_autorita

- tra i nomi di tipo 'R': *Brunico; *Archivio di stato di Bologna; *Biblioteca *Capitolare <Benevento>, Biblioteca *Palatina <Firenze ; 1844>; *Italia <Regno : *Ministero delle *Finanze>;
- tra i nomi 'G' ad es.: *Minori *conventuali : *Convento di *Santa Maria di Betlem <Sassari>, *Torcello, *Chierici *Regolari
- nell'analisi a campione si è rilevato che circa la metà dei nomi di autori identificati di MOL sono già presenti in SBN spesso in forme diverse.

Nell'evoluzione di MOL si prevede di uniformare il formato dei dati a SBN.

Conclusioni

I dati di autorità sono fondamentali per l'accesso alle risorse e quindi la loro coerenza e correttezza per un'interfaccia di ricerca integrata è un requisito imprescindibile.

Perché il lavoro di clusterizzazione dei dati di autorità di nomi e di luoghi possa realizzarsi occorre che ogni singolo dato venga esaminato e confrontato. Oltre a servirsi di sistemi automatizzati per il rilevamento di uguali e di simili non si può fare a meno di intervenire manualmente. Considerando la quantità di dati presenti nei tre archivi, il gruppo valuta indispensabile la programmazione di un sotto-progetto specifico.

Cfr. Allegato 9):

[SRI_DLG_CTecnico_tabella_confronto_nomi_MOL-SBN](#)