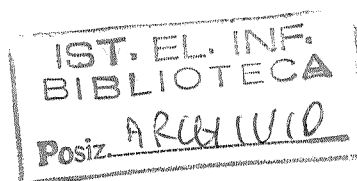


Consiglio Nazionale delle Ricerche

**ISTITUTO DI ELABORAZIONE
DELLA INFORMAZIONE**

PISA



**Normalizzazione dei dati bibliografici per
la costruzione del catalogo collettivo della
Biblioteca di Area: ATLASFL - ISO 2709**

Roberta Amaranti, Stefania Biagioni

Nota interna B4 -14

Giugno 1994

Descrizione della testata:

POSIZIONE	LUNGHEZZA	DESCRIZIONE
0	5	lunghezza record
5	1	stato del record
6	4	codici di implementazione
10	1	lunghezza degli indicatori
11	1	lunghezza degli identificatori di sottocampo
12	5	indirizzo relativo di inizio dei dati
17	3	ad uso del sistema
20	1	lunghezza del secondo sottocampo del campo indice
21	1	lunghezza del terzo sottocampo del campo indice
22	2	per usi futuri

1

¹ Errata corrige pag 2



CONSIGLIO NAZIONALE DELLE RICERCHE
ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

**Normalizzazione dei dati bibliografici
per la costruzione del catalogo collettivo della
Biblioteca di Area: ATLASFL - ISO 2709**

Roberta Amaranti
Stefania Biagioni

INTRODUZIONE.....	1
1. Il formato ATLAS - FL.....	2
2. Formato di interscambio ISO 2709	2
2.1 Formato degli archivi ISO accettati da CDS/ISIS	5
3. Descrizione dei programmi e analisi delle operazioni	6
3.1 Convert.....	6
3.2 Fliso.....	6
3.3 Fangorn	6
4. I programmi ed il loro uso.....	7

INTRODUZIONE

Il presente lavoro si colloca nell'ambito delle attività svolte dal Gruppo di lavoro Biblioteca di Area di Pisa nel progetto del catalogo collettivo della Biblioteca di Area. Il Gruppo ha infatti deciso di reperire le registrazioni bibliografiche già presenti nelle biblioteche degli istituti del CNR di Pisa, e di convertirle nel formato standard ISO 2709 per la successiva immissione nel sistema CDS/ISIS prodotto dall'Unesco adottato per la realizzazione del Catalogo Collettivo.

La prima fase del lavoro ha riguardato l'analisi delle strutture dei dati prodotti dai singoli sistemi presenti nei vari istituti (quando presenti).

Uno dei sistemi è Atlas nella sua versione di registrazione "fuori linea", ed è proprio della descrizione dei programmi di conversione di queste registrazioni nel formato ISO 2709 che il presente lavoro si occuperà presentando brevemente la descrizione dei singoli formati, dei programmi necessari alla conversione e l'uso del software nel contesto del procedimento complessivo di conversione.

I dati vengono convertiti, quindi, dal formato Atlas "fuori linea" in un formato intermedio che subirà alcune modifiche per diventare conforme al formato ISO 2709 richiesto dal sistema CDS/ISIS.

1. Il formato ATLAS - FL

E' il record prodotto dal programma di registrazione "fuori linea" dei dati importabili nel sistema ATLAS in forma codificata ¹ .

2. Formato di interscambio ISO 2709

Il formato ISO 2709 è un formato standard internazionale per lo scambio delle informazioni raccomandato dall' ISO (International Standard Organization).

Il record ISO è formato da tre parti:

Testata: contiene elementi di lunghezza fissa in cui sono registrate le informazioni necessarie alla elaborazione del record.

Indice: contiene campi a lunghezza fissa, che identificano rispettivamente: l'etichetta, la lunghezza e la posizione dei vari campi dati nel record. L'indice termina con un separatore di campo.

Campi dati: contiene informazioni alfanumeriche a lunghezza variabile, separate da un codice di fine campo.

Le informazioni vere e proprie sono contenute nella terza parte del record ed è possibile estrarle grazie alle informazioni contenute nella testata e nell'indice. Ogni record è a lunghezza variabile ed un separatore ne segnala la fine.

Descrizione della testata:

POSIZIONE	LUNGHEZZA	DESCRIZIONE
0	5	lunghezza record
@ C1 :	Collezione	
@ C3 :	Sottocollezione	
@ C5 :	Titolo originale	

¹ ATLAS - Manuale per la registrazione fuori linea dei dati bibliografici - M.B. Baldacci, O.G.Mancino, R.Sprugnoli

@ C6 :	Titolo convenzionale
@ D1 :	Note
@ E1 :	Anno di pubblicazione
@ E3 :	Paese di pubblicazione
@ E4 :	Lingua
@ E5 :	ISBN
	Autore principale
@ F2 :	Autore secondario
@ G2 :	Ente principale
@ G2 :	Ente secondario
@ G3 :	Classificazione
@ I1 :	Soggetto autore
@ I1 :	Altro soggetto autore
@ H1 :	Soggetto
@ H1 :	Altro soggetto
@ J1 :	Collocazione
@ J2 :	Inventario

1) Lunghezza del record (5 caratteri), e' un valore numerico decimale che riporta la lunghezza dell'intero record.

2) Stato del record (1 carattere), e' significativo solamente negli archivi di esportazione e assume il valore "0".

3) Codici di implementazione (4 caratteri)

Sono codici definiti per ogni specifica implementazione dello standard ISO 2709 (ISISXCH pone questo campo a '0000' quando produce un file in uscita).

d. Lunghezza degli indicatori (1 carattere)

In questo campo viene indicata la lunghezza degli indicatori che compaiono all' inizio di ogni campo. Tali indicatori forniscono informazioni sul contenuto del campo.

e. Lunghezza degli identificatori di sottocampo (1 carattere)

In questo campo viene indicata la lunghezza degli identificatori di sottocampo (solitamente due caratteri).

f. Indirizzo di inizio dati (5 caratteri)

Il contenuto di questo campo indica il numero dei caratteri di controllo (lunghezza della testata piu' lunghezza dell'indice e separatore di campo) che precedono i campi a lunghezza variabile. Se questo numero viene sommato all'indirizzo relativo di inizio dati di un certo campo si ottiene l'indirizzo del primo carattere di quel campo.

g. Ad uso del sistema (3 caratteri)

Questa posizione puo' contenere dati legati al sistema informativo che produce l'archivio. Viene posto a '000' negli archivi in uscita prodotti da CDS/ISIS.

h. Lunghezza del secondo sottocampo del campo indice (1 carattere)

Questa posizione indica la lunghezza (numero decimale) del secondo sottocampo del campo indice, cioè del sottocampo in cui è definita la lunghezza del campo dati a cui si riferisce.

i. Lunghezza del terzo sottocampo del campo indice (1 carattere)

Questa posizione indica la lunghezza (numero decimale) del terzo sottocampo del campo indice, cioè del sottocampo che indica la posizione del primo carattere del campo relativamente al primo carattere di inizio dati.

j. Per usi futuri (2 caratteri)

Contiene '000'.

Indice

Ogni campo dell'indice è di lunghezza fissa ed è costituito da tre sottocampi: il primo contiene una etichetta o tag che identifica il campo; il secondo indica la lunghezza del campo; il terzo indica la posizione del primo carattere del campo riferita all'inizio della parte dati del record.

La lunghezza di un campo dell'indice varia a seconda dell'archivio, ma può essere identificata nelle posizioni 20 e 21 della testata del record, che indicano la lunghezza rispettiva del secondo e terzo sottocampo.

a. Etichetta (3 caratteri)

Questo è un numero decimale che identifica il campo. I valori delle etichette sono numeri decimali. Il significato delle etichette dipende dal programma che eseguirà l'elaborazione dei dati.

b. Lunghezza del campo

Indica la lunghezza (in caratteri) del campo dati corrispondente ed include il separatore di campo.

c. Posizione del campo

Indica la posizione del primo carattere del campo rispetto al primo carattere del campo dati del record. Sommando questo valore a quello dell'indirizzo di inizio dei dati si ottiene l'indirizzo assoluto del primo carattere del campo.

Campi dati di lunghezza variabile

Dopo la testata e l'indice, il record contiene i campi dati.

Per ogni campo dati del record c'è un campo nell'indice che gli corrisponde e per individuare la posizione di inizio di un dato campo è necessario ricercare nell'indice l'etichetta corrispondente al campo desiderato. I campi dati di lunghezza variabile sono disposti nel record nella stessa sequenza dei corrispondenti campi nell'indice. Sia all'indice che ad ogni campo di lunghezza variabile fanno seguito i separatori di campo. Il record finisce con un separatore di record.

2.1 Formato degli archivi ISO accettati da CDS/ISIS

Gli archivi ISO prodotti da CDS/ISIS sono di tipo testo nel formato descritto in precedenza. Poichè è abbastanza scomodo manipolare archivi testo con linee relativamente lunghe, specialmente quando è necessario ispezionarle con un editor o trasmetterle su una linea di comunicazione, CDS/ISIS spezza il record in segmenti di 80 caratteri, ognuno dei quali è seguito dai caratteri di fine riga standard ^M ^J (ritorno carrello e avanzamento linea). Tutti i segmenti, eccetto l'ultimo, conterranno esattamente 80 caratteri di testo.

3. Descrizione dei programmi e analisi delle operazioni

3.1 Convert

Questo programma² serve per decodificare i dati dal formato ATLAS - FL in un formato testo facilmente leggibile.

3.2 Fliso

Questo programma³ effettua operazioni di normalizzazione, in base alle specifiche stabilite all'interno del Gruppo di lavoro della Biblioteca di Area, sui dati bibliografici presentati in formato testo dal programma Convert, restituendo in uscita un file pronto per essere convertito in formato ISO 2709.

3.3 Fangorn

Questo programma⁴ (che attualmente gira sotto MS/DOS, PC/DOS, OS/2) è nato dalla necessita' di recuperare dati registrati in formato ASCII in uscita da un qualsiasi sistema e ristrutturarli opportunamente secondo le specifiche richieste dal sistema ISIS (cioè conformi al formato ISO 2709). Tale operazione puo' essere fatta solo se nelle registrazioni sono individuabili i campi tramite particolari caratteristiche formali.

Il file di origine puo' essere convertito interamente oppure si possono specificare gli elementi che si vogliono trasportare nel nuovo data base. A tale scopo deve essere preparato un file di specifiche (ASCII) scritto con un elaboratore di testi seguendo una serie di regole. Dopo che le specifiche sono state lette con successo e memorizzate viene chiesto di specificare il nome del file in ingresso ed il nome di un file ISO 2709 in uscita.

² G. Sbrana - Universita' degli studi di Pisa

³ R. Amaranti - Consulente informatica

⁴ H. Besemer, P. Nieuwenhuysen: Fangorn Manuale d'uso

Dopo che i nomi sono stati specificati il programma inizia ad elaborare i dati visualizzando il numero del record che sta elaborando. Al termine del lavoro il programma emette un segnale acustico.

4. I programmi ed il loro uso

Software utilizzato: Atlas_fl, Fangorn, Convert, Norm, Fliso.

Supponendo di avere tutto il software necessario memorizzato in un'unica directory di nome ATLAS la prima operazione da compiere è sottoporre i dati in formato Atlas_fl (forniti su un floppy disk nell'unita' a:) al programma **Convert** che elabora i dati e permette, quando è il caso, di eliminare i record non utilizzabili per la conversione (record corrotti).

Per mandare in esecuzione il programma **Convert** basta digitare all'interno della directory ATLAS il comando *convert*.

Il programma va in esecuzione visualizzando sullo schermo i numeri dei record che sta elaborando.

Le possibili cause di interruzione del programma sono le seguenti:

1) Termina con successo visualizzando un messaggio di fine elaborazione: in questo caso i record presenti sul dischetto sono stati esaminati tutti senza riscontrare anomalie.

2) Termina bloccandosi sul record corrotto visualizzando il numero progressivo.

In questo caso si lancia il programma Atlas (viene richiesto il codice della biblioteca che ha registrato i dati), si seleziona dal menu' principale la voce **5) Visualizzazione** e si ha la possibilità di vedere qual'è il record sbagliato memorizzando il suo numero (per

uscire dalla visualizzazione del record il comando è : **system**). Successivamente, dopo essere ritornati al menu' principale di Atlas, si seleziona la voce **4) Eliminazione** per cancellare il record errato; fatto questo deve essere riattivata l'esecuzione del programma **Convert** per continuare l'elaborazione dei dati.

N.B. : alla fine della elaborazione il file che viene generato nella directory ATLAS ha il nome **Filseq**.

A questo punto, se il programma **Convert** è terminato con successo, deve essere mandato in esecuzione il programma **Norm** che prende in ingresso il file **Filseq**, generato da **Convert**, e produce in uscita un file di nome **Filis**. Per far eseguire quest'ultimo programma è sufficiente digitare all'interno della directory ATLAS il comando *norm*.

Il passo successivo, dopo che è stato eseguito il programma **Norm** con successo, consiste nel mandare in esecuzione il programma **Fliso** che prende in ingresso il file di nome **Filis** generato nella fase precedente e restituisce un file di nome **Mut**.

Per far girare il programma **Fliso** è sufficiente digitare, all'interno della directory ATLAS il comando *fliso*.

Esempio di file Filis in ingresso al programma Fliso

Ogni campo del file e' preceduto da una etichetta del campo contenente il carattere "@" ed una sigla seguita da ":".

ETICHETTA	DESCRIZIONE
@ A1 :	Titolo generale
@ C1 :	Collezione
@ C3 :	Sottocollezione
@ C5 :	Titolo originale
@ C6 :	Titolo convenzionale
@ D1 :	Note
@ E1 :	Anno di pubblicazione
@ E3 :	Paese di pubblicazione
@ E4 :	Lingua
@ E5 :	ISBN
@ F1 :	Autore principale
@ F2 :	Autore secondario
@ G2 :	Ente principale
@ G2 :	Ente secondario
@ G3 :	Classificazione
@ I1 :	Soggetto autore
@ I1 :	Altro soggetto autore
@ H1 :	Soggetto
@ H1 :	Altro soggetto
@ J1 :	Collocazione
@ J2 :	Inventario

N.B.: il file **Mut** prodotto in questa fase non è ancora pronto per essere tradotto nel formato ISO 2709, ma necessita di alcune modifiche sul testo da effettuare manualmente se in esso sono contenute registrazioni di opere in piu' volumi.

Trattamento delle opere in piu' volumi

Le istruzioni che seguono indicano i passi per trasformare la descrizione a due livelli di un'opera in piu' volumi (livello generale e livello monografico) nelle corrispondenti descrizioni di livello monografico.

Ogni opera in piu' volumi è caratterizzata da una intestazione la cui prima riga, come la prima riga di ogni altro record, è etichettata con "@ **: ***" , mentre ogni singolo volume è distinto da una riga etichettata con "@ **: +++". Le operazioni manuali da effettuare (con un qualsiasi editor) consistono in:

- copiare il campo etichettato con "@BB:" dalla descrizione generale alla descrizione di ogni singolo volume subito dopo la prima riga (quella etichettata con "@ **: +++");
- copiare il campo etichettato con "@E4:" dalla descrizione generale alla descrizione di ogni singolo volume nella posizione data dall'ordinamento (le etichette sono infatti ordinate alfabeticamente);
- controllare se il campo etichettato con "@A1:" della descrizione generale (campo titolo generale) coincide con i campi etichettati con "@A1:" dei singoli volumi (campo titolo del singolo volume); in caso affermativo non è necessario ripeterlo, ma basta aggiungere, nel campo titolo di ogni singolo volume, il corrispondente numero del volume (ad es. vol.1 ecc.); in caso contrario l'etichetta "@A1:" (ISBD) dell'intestazione viene cambiata in "@C7:" e viene copiata questa nuova riga dell'intestazione così modificata in ogni singolo volume aggiungendo il corrispondente numero di volume;
- cancellare l'intestazione e sostituire i "+++" con "***" in ogni prima riga dei singoli volumi.

L'ultima modifica manuale consiste nel sostituire nel testo del file un asterisco dopo un articolo con parentesi angolate, ad esempio "the * cluster " diventa "<the >cluster" e "Il * codice genetico" diventa "<Il >codice genetico".

A questo punto il file **Mut** è pronto per essere elaborato con il programma **Fangorn**.

Unitamente al testo deve essere passato al programma Fangorn un file di specifiche di nome **Atl.spe**. Il file che Fangorn, nel nostro caso, produce è **Iso**. Il programma Fangorn viene attivato digitando, all'interno della directory ATLAS, il comando `fangorn`. Di seguito, con un editor è possibile visualizzare il file prodotto in uscita da Fangorn e verificarne la correttezza.

Il passo successivo è quello di attivare la funzione di importazione del programma ISIS30 che richiede i seguenti dati:

Nome del file di entrata: Iso.iso;

Separatore di campo: ~; (oppure "ALT 126" se la tastiera è italiana)

Separatore di record: #;

FST di riformattazione: impo.fst;

Tabella di conversione: mutam.par.