

Data Magician®

Un software per la riformattazione dei dati bibliografici e la conversione dei cataloghi

di Francesco Dell'Orso

Chi legge troverà qui la rassegna analitica e ragionata di un programma, Data Magician, per la manipolazione di record bibliografici volta a mutarne il formato, e non troverà pertanto un'introduzione alla descrizione della riformattazione in generale, per cui rimando ai riferimenti in nota.¹ Tuttavia, qualcosa di generale ed introduttivo verrà detto, e inoltre, viste le caratteristiche del prodotto esaminato, l'esame del particolare richiederà le istanze generali.

A cosa serve la riformattazione dei dati bibliografici? Sostanzialmente a due cose. A caricare automaticamente in un qualche sistema per gestire database (Dbms-Database management system) dei dati già esistenti e scritti su supporto magnetico, e quindi a poterli riutilizzare senza doverli reimmettere da tastiera. A convertire, trasformare i dati per quanto attiene a formato e/o contenuto.

Il formato strutturale dei dati va mutato per farli migrare fra software che ne impongono uno diverso, e per potere così, ad esempio, tramutare file Dbf gestiti con dBase in dati trattati da ProCite, transitando per il formato di scambio con

delimitatori detto *comma delimited*, oppure per passare da un formato ad etichette all'Iso 2709, o da un file scritto con Ms-Word ad un Dbms, per la passerella di un file di testo Ascii.

Vari sono gli elementi di struttura per cui i formati possono essere differenti e incompatibili: per gli elementi che caratterizzano inizio e fine di record e di campo, per lunghezza e posizione degli elementi, per il tipo di registro che tiene traccia degli estremi di queste dimensioni, per segnali di controllo inseriti nel contenuto, e perché quasi ogni produttore codifica infine i file in formato proprio, appunto "proprietario" non immediatamente leggibile da un altro software. Il contenuto dei record viene riformattato per cambiare posto, nome, forma e contenuto ai dati: riconoscere per unificare o separare, omettere o contrassegnare distintamente, ad es., luogo, editore e data; titolo, numero di volume e di fascicolo, data e pagine della rivista in cui è stato pubblicato un articolo; nome, cognome, date, titoli di un autore ecc.; si può volere alterare il contenuto dei record e la disposizione degli elementi anche

all'interno dello stesso sistema, autotrasfondendo i dati per modificarli e conservarli.

Perché e quando riutilizzare o trasformare i dati? Cataloghi informatizzati di biblioteca, database bibliografici e testuali, archivi su pc gestiti con Dbms, dati riportati dalla carta (schede, fotocopie di indici ecc.) in formato elettronico grazie alla lettura ottica, file scritti con un word processor... sia che si trovino on line, in Internet o no, off line, su dischetto o sul disco rigido di un pc, su cd-rom, tutti veicolano materiale informativo esistente, che può circolare nelle mani di più di un utilizzatore e in più di un sistema software di gestione: "fatti voi foste per esser adoprati, non per blandire chi vi ha creati", così il Nostrò. Dei sistemi bibliotecari possono desiderare di fondere database diversi in uno unico omogeneo,² un individuo può volersi garantire la possibilità di passare in futuro all'uso di un altro software, senza subire la manomorta di alcun prodotto. In certi ambienti è normale che chi studia legga quanto scritto da altri, per questo faccia ricerche bibliografiche in archivi informatizzati — appunto opac locale, cd-rom, Internet ecc. —, trovi riferimenti, anche con abstract, anche molti, e intenda servirsi di quelle citazioni — per scrivere un articolo o un libro o tenerle nel proprio schedario su personal computer —; accade che costui preferisca copiare elettronicamente quei dati — scaricarli e ricaricarli con o senza modifica: downloading-reformatting-uploading — piuttosto che ricopiarli, una volta, se ha almeno una stampa dei dati in hardcopy, o due volte, se non ce l'ha.³ Altrove dei bibliotecari vogliono utilizzare i dati presenti nel catalogo centrale in linea per allestire database, cataloghi o elenchi particolari, per darli o scambiarli con altre biblioteche e persone. Sembrerebbe che la gestione delle citazioni bibliografiche

possa non arrestarsi all'interno dei confini della gestione bibliotecaria, pure sensibile alla ripetizione delle operazioni di elaborazione ed output sullo stesso nucleo di dati bibliografici (ordinazione, fatturazione, inventariazione, catalogazione diversa, prestito ecc. ecc.) tanto che questa giustificò e fomentò l'introduzione dell'automazione in biblioteca. Anche l'informazione bibliografica come unità fisica dei

dati può circolare quando il sistema centrale, e chi lo gestisce, si pongono come veri "serventi" (server), nei confronti degli utilizzatori: assumere e tenere tale posizione costituisce un lavoro di onesta e modesta intelligenza tecnica ed organizzativa, in grado poi di evitare innumerevoli duplicazioni di operazioni, provocando risparmio di tempo e denaro, il che per me spiega perché in molti altri am-

bienti le cose funzionino diversamente.

Disponendo di un file scritto, poniamo con Wordstar 4 o WordPerfect 6 e dovendo lavorare in Ms-Word, penseremo a ricopiare il testo a mano o a convertirlo? I word processor oggi per lo più hanno incorporate queste funzioni di conversione, che non prevedono intervento dell'operatore, ma esistono anche software specifici, dei ►

Tab. 1 - Funzioni e comandi di Data Magician in sommario

- aggiungere stringhe di caratteri —all'inizio, in fondo, dove si vuole
- porre condizioni —rappresentate da stringhe presenti/assenti, da assenza/presenza di campo o sottocampo o occorrenza, dall'uguaglianza di un valore (alfabetico o anche numerico: numero di record, data...), dal confronto con un'altra condizione, vicina o lontana, già posta e così via
- usare operatori booleani —and or not
- cercare stringhe di caratteri ed anche rimpiazzarle con altre
- cercare stringhe ed estrarre —quindi frammentare ed enucleare: *parsing*— ossia tagliare un campo ed estrarne del contenuto, vagliando condizioni, assumendo come punto di incisione: una qualsiasi stringa di caratteri in esso contenuta (caratteri jolly compresi); un carattere in una data posizione; una parola in una data posizione —numero sequenziale nella "frase"; un'occorrenza del campo indicata per numero di sequenza; il delimitatore delle occorrenze interne ad uno stesso campo —potendo poi sempre anche procedere a destra o sinistra secondo quanto indicato
- aggiungere, eliminare, fondere campi diversi e porzioni del contenuto di essi
- convertire maiuscolo e minuscolo in vario modo
- eliminare punteggiatura e spazi in luoghi e modi diversi
- trattare la data
- impostare e incrementare un contatore numerico
- usare un file esterno per controllare e convertire stringhe (*look up table*)
- trattare caratteri jolly, *wildcard*, come i tradizionali ? e *, e poterli anche ridefinire con altri, poterli usare nel cerca/rimpiazza
- trattare tutto il set esteso dei caratteri ASCII

I vari comandi possono essere combinati fra loro, ossia posti in sequenza, anche con quelli condizionali, ad es. tre record in input presentano un campo come:

```
IND: 06123 214A CA, Ontario
IND: 0123456789 IT, Toscana
IND: A34 00 7Z GW, Mannheim
```

si può impostare un'istruzione che prescrive: "*cerca la stringa 'IT, ' e se c'è scrivi in output solo quanto la segue premettendo 'Italia', fa lo stesso con 'CA, ' per il Canada, altrimenti scrivi solamente che 'Non è né Canada né Italia'*"

i comandi sono:

```
c" IT, " BR" IT, " TB"Italia "C~ CS1 C" CA, " BR" CA, " TB"Canada " C~ Cs2 Cr1
AR2 XC "Non è né Canada né Italia"
```

l'output relativo a quei tre record è:

```
Canada Ontario
Italia Toscana
Non è né Canada né Italia
```

convertitori puri, quali Word for Word della Mastersoft.⁴ I software di gestione di archivi bibliografici conoscono qualcosa di molto simile: vari fra essi sono corredati di un programma di riformattazione gregario — incorporato come proprio o distinto perché aggiunto — che serve a trattare i dati scaricati da varie fonti ed in diversi formati (*n* input) per convertirli nel formato proprietario del database ricevente (*1* output): Procite ha BiblioLink, Reference Manager ha Capture, micro Cds/Isis ha Fangorn, EndNote ha EndLink, Texto ha Textload, Papyrus, Library Master, Idealist hanno la funzione incorporata. Spesso operano in modo automatico, predisposti dai produttori a trattare i dati dei database più importanti, ma consentono anche la personalizzazione dei filtri di riformattazione. La varietà di fonti non è estranea alla non immediatezza dell'operazione, perché, co-

me ricorda Sue Stigleman, scaricare i dati è semplice, importarli no, non c'è ancora un formato standard per i dati che vengono esportati da database bibliografici per venire importati in archivi personali su pc. Il sistema Melvyl è stato fra i primi sensibili al problema instaurando la collaborazione fra biblioteche e case di produzione (Niles, Balboa, ex-Pbs, Ris e Rsd).⁵ La riformattazione dei dati opera normalmente su una copia dei dati, lasciando intonsa la fonte di essi; per una descrizione delle fasi operative di cattura dei dati, scaricamento, eventuale ripulitura preventiva,⁶ analisi ecc. rimando ancora a quanto citato nella nota 1. Comunque, prima di cominciare a por mano alla manipolazione dei dati, sarà il momento di ripassare alcuni precetti:

a) analizzare e conoscere i dati in input in dettaglio, anche su una stampa;

b) conoscere l'articolazione dei campi nei record del database di destinazione;

c) procedere dal generale al particolare, dal semplice all'articolato: record, campo, contenuto, ecc.;

d) aggiungere comandi progressivamente;

e) provare inizialmente su pochi record;

f) ricordare — a mo' di Legge di Murphy — che, come molte altre cose che hanno a che fare con le biblioteche e coi computer, anche non assieme: sarà meno facile di quello che sembra, richiederà più tempo di quanto stimato.

Data Magician non è un software puro di manipolazione testuale (come Awk, Snobol, il *macrolanguage* di Kedit ecc.), ma un convertitore per la riformattazione di file di record bibliografici. È uno strumento che sta in mezzo, fra i dati provenienti da fonti varie e le varie destinazioni che essi potranno ave-

Tab. 2 - Caratteristiche analitiche

1. FILE: ci sono comandi a livello dell'intero file	Sì, sia comandi che definiscono il formato del file (<i>delimited</i> , <i>tagged</i> , MARC, ISO 2709) includendo separatori etc., sia comandi (<i>pre-processing</i>) che trattano preventivamente ogni linea di input prima della manipolazione interna
2. FILE: il formato può venire automaticamente riconosciuto	no
3. RECORD: dimensioni massime	10.000/12.000 caratteri
4. RECORD: come riconoscere l'inizio (a) in base ad una posizione fissa (inizio linea, nuova riga come nei <i>delimited</i>) (b) in base ad un'etichetta individuata per: (b1) posizione, fissa o variabile e/o (b2) contenuto (come "Tipo di record", "Intestazione"...)	riconosce automaticamente i formati: dBase, INMAGIC, STAR, ISO 2709, MARC; per i <i>delimited</i> occorre indicare il numero dei campi, i delimitatori e i separatori; per i <i>tagged</i> l'etichetta può essere in posizione fissa o variabile ("L'inizio contiene...")
5. RECORD: come riconoscere la fine (esplicita o l'inizio del successivo)	l'inizio del record successivo basta come fine del precedente
6. RECORD: inizio e continuazione possono venire letti e riconosciuti oltre e attraverso scritte da trascurare (ad es. il record si espande oltre uno schermo/pagina, ci sono: <i>segue...</i> , messaggi, intestazioni, numerazione...)	sì, può trattarle <i>tout court</i> come etichette da non assumere o può trattarle e scartarle preventivamente nei <i>pre-processing codes</i> — cfr. 1.
7. RECORD: si possono eliminare/abbandonare uno o più record durante la fase di conversione	sì, in base alla verifica di una o più condizioni ("SE...ALLORA...")

re, non è incorporato in alcun sistema di gestione di dati. È in grado di leggere, manipolare, ricreare la struttura, il formato vero e proprio di un file: ad es., leader e directory, delimitatori di fine campo e fine record del formato Marc. È predisposto per riconoscere e trattare automaticamente in input ed in output vari tipi di file: con delimitatori (*delimited*), con etichette (*tagged*), in formato Marc, "senza formato" (cfr. oltre per queste virgolette), in modo da prendere e fornire dati a: dBase, Inmagic, Marc, Star, Micro Cds/Isis, Library Master. Dal momento che può trattare file nei formati *passee-partout* "con delimitatori" e "ad etichette" e file "privi di formato", si avvicina ad essere uno strumento aperto per la manipolazione di struttura e/o contenuto di dati testuali, inclusi Sgml e derivati, Tei o Html. Data Magician offre una batteria di comandi, vasta e potente per alte-

rare il contenuto dei vari campi. Dato che i vari comandi possono venire uniti congiuntamente ed intersecati come componenti di un linguaggio, la riformattazione che ne scaturisce è straordinariamente versatile, limitata soprattutto dall'esperienza e fantasia — creativa, sì, è vero — di chi se ne serve. Esso dà modo di avere ragione dell'eccentrica strutturazione che i dati possono avere in campi di cruciale importanza per i record bibliografici come quelli dei nomi degli autori, dell'area della pubblicazione, del titolo del documento ospite per le parti componenti. Nessuno dei programmi prima citati può occupare una posizione vicina a quella di Data Magician; considerandoli da vicino è più corretto dire che essi sono strutturalmente non omogenei. Per un confronto fra simili occorre piuttosto rivolgersi a prodotti come Headform, Refwriter, entrambi inglesi, e, forse, all'i-

taliano Malachia.⁷ Una sfida alla riformattazione e, a quanto sembra, una campagna in corso, anche in Italia, è quella offerta dai record "senza formato", *unformatted*. In realtà, qualsiasi programma, di riformattazione o no, riesce a funzionare proprio solo e in quanto sa rinvenire delle forme, delle costanti, delle regole, cui applicare algoritmi, inferenze ed azioni. Quindi anche questi record che si definiscono senza formato e struttura, ne hanno pure uno: certo, schede di catalogo — in formato ISBD o no — non sono una tisana per i software di riformattazione rispetto a quelli che identificano i campi per posizione o per nomi univoci. È ufficiale che "Lo scopo principale dell'ISBD è facilitare la comunicazione internazionale delle informazioni bibliografiche... (iii) favorendo la conversione delle registrazioni bibliografiche in forma leggibile dal- ➤

8. RECORD: c'è un contatore, lo si può impostare, controllare e porne il valore in un campo	sì, ad es.: NS 170 NU NI10 imposta il contatore a 170, registra il valore nel campo corrente, incrementa di 10 (180, 190...) per ogni record
9. CAMPO: numero totale di campi ammessi	non c'è limite dichiarato (> 200)
10. CAMPO: come riconoscerlo nei formati ad etichette (negli altri formati lo fa automaticamente) (a1) etichetta a posizione fissa o variabile (a2) etichetta a lunghezza fissa o variabile (a3) includente caratteri speciali, sotto ASCII 32 (a4) con necessario separatore dal campo (a5) a struttura fissa	(a1) l'etichetta può anche essere a posizione non fissa e (a2) a lunghezza fissa o variabile; (a3) può trattare ASCII <32 come, ad es. <015>, (a4) il separatore non è indispensabile, ma aiuta, (a5) la struttura è comune a tutte le etichette
11. CAMPO: come riconoscere la fine del campo	dall'etichetta che segna l'inizio del successivo
12. CAMPO: (a) il contenuto può continuare sulle linee successive; (b) è necessario indicare come	(a) sì; (b) non indispensabile, ma consigliabile
13. CAMPO: ordine diverso in record diversi	sì, a meno che non siano <i>delimited</i>
14. CAMPO: cambiarne il numero/quantità (a) eliminare/abbandonare uno o più campi durante la conversione (b) fondere più in uno (c) duplicare (d) aggiungere ex-novo	(a) sì, non fornendogli destinazione ed anche in base alla verifica di una o più condizioni (b) sì (c) sì (d) sì

la macchina",⁸ e l'articolata sintassi della sua prosa genera descrizioni dei documenti rappresentative ed identificative di essi, intellegibili all'occhio umano che legge i dati nel contesto dell'intera descrizione. Ma l'approccio della maggior parte dei software di riformattazione è ben meno globalmente intelligente di quell'occhio e si scontra subito con difficoltà non secondarie dello stile ISBD, in cui, pur a prescindere dai vari livelli di completezza della catalogazione: i campi (aree o elementi che siano) non hanno nome, solo le aree hanno un distintivo iniziale. " - ", ma i capoversi, anche senza rientri, possono alterarlo; il numero delle aree è fisso solo come massimo tipologico, perché le aree possono mancare (il dato non c'è o una ISBD non lo prevede); all'interno delle aree, gli elementi hanno scivolosità ancora maggiore: ripetizione di uno o più,

assimilazione per legame grammaticale, punteggiatura convenzionale — quasi mai appartenente ad una unica area — confondibile con quella di contenuto. Data Magician può trattare dati "senza formato", da record ISBD a citazioni di stile bastardo (un esempio nella Tab. 4), alle voci di un indice di una rivista: naturalmente, accontentarsi di un basso grado di analicità, o avere a che fare con record "semplici" rende l'impresa meno ardua. Tuttavia, è da annotare che Data Magician dispone di fino a 99 "segnalini" in cui memorizzare temporaneamente un valore di verità, conseguente a un controllo, in modo da potere stendere una rete, anche fitta, di "se... allora" per effettuare altri controlli e, in ragione di questi, eseguire azioni. Realizzare "un prototipo capace di scansionare le schede di catalogo [ISBD e non - nda] e di formattare i dati in

Unimarc" sembra comunque un'altra impresa. È infatti l'impresa del progetto Facit, meritevole di un finanziamento comunitario europeo, con un costo di oltre un miliardo di lire italiane.⁹

Data Magician è un programma interamente Dos, a semplici menu, e che funziona anche sotto Windows. Arrivato alla versione 1.4 release 3[.9], è l'opera di Lawrence Folland, canadese del grande bianco nord.¹⁰ Offre un help circostanziato (*context sensitive*), ottimo dal punto di vista informativo, anche con indice, e schermi numerati cui si può saltare direttamente. Per scrivere i comandi, si dispone di un decoroso strumento di scrittura che accetta istruzioni indifferentemente in minuscolo/maiuscolo, liberamente separate da spazi, e che fa apporre ovunque commenti preceduti da "!" ; comprende comandi di "Undo", di distruzione rapida, e per il lavo-

Tab. 2 - Caratteristiche analitiche (segue)

<p>15. SOTTOCAMPI: possono venire: (a) trattati; (b) distinti dalle occorrenze di uno stesso campo ripetibile —cfr. 16.; (c) modificati quanto agli identificatori; (d) mutati di posizione all'interno del campo; (e) eliminati; (f) fusi; (g) duplicati; (h) aggiunti ex-novo</p>	<p>sì, tutto</p>
<p>16. OCCORRENZE (ripetizioni di uno stesso campo): possono venire: (a) riconosciute; (b) avere delimitatore diverso nei vari campi; (c) identificate e trattate in base al numero di sequenza; (d) fuse; (e) create</p>	<p>sì, tutto</p>
<p>17. STRINGHE DI CARATTERI ("testo"): all'interno del campo possono venire: (a) riconosciute; (b) soppresse; (c) sostituite; (d) manipolate; (e) create; (f) sulla base di: (f1) posizione fissa del carattere d'inizio, e di fine (f2) numero di parola; (f3) presenza/assenza di quelle o altre stringhe di caratteri</p>	<p>sì, tutto, senza limitazioni</p>
<p>18. STRINGHE DI CARATTERI: si possono trattare: (a) spazi (ASCII 32) ridondanti (b) punteggiatura (c) set ASCII esteso inclusi caratteri speciali (< ASCII 32) (d) minuscolo/maiuscolo (e) caratteri jolly (<i>wildcard</i>)</p>	<p>(a) sì, c'è anche comando per eliminarli all'inizio, in mezzo, alla fine del campo (b) <i>idem</i>, specificandola o per <i>default</i> (c) sì, ovunque, come, ad es. <015> (d) sì, convertire tutto, o solo l'inizio della prima o di tutte le parole (e) sì, anche ridefinibili, si possono cercare e sostituire</p>
<p>19. CONTROLLO DI VALIDAZIONE con dati esterni</p>	<p>sì, "look-up table" (limitato)</p>

ro sui blocchi.

L'ambiente operativo è confortato da altre funzioni sempre disponibili: l'uscita temporanea al Dos; la visualizzazione a sommario di tutti i campi definiti in input; quella ancor più importante dei file esterni di input e di output; la stampa, anche su disco, delle impostazioni di riformattazione; il dimensionamento, sullo schermo, di righe e finestre di lavoro per aggiungere o togliere linee, campi, contenuto dei campi, codici, ecc.; stretto controllo, e non solo sintattico, sulla correttezza dei comandi di manipolazione.¹¹ Si può trattare, sia in input che in output, un record alla volta, distinguendo lettura e riformattazione dei dati: in questo modo si vede cosa viene assunto e, con un successivo comando, cosa producono i comandi su di esso: è la riformattazione in tempo reale sotto lo zoom.¹² Così si possono provare

le istruzioni di elaborazione progressivamente e in relativa interattività (relativa perché durante la conversione non si può accedere ai dati fisici per correggerli, ma si può interromperla, correggere le istruzioni di formattazione, procedere o ricominciare). L'esecuzione del procedimento può anche svolgersi in un'unica soluzione e venire osservata, mentre viene anche mostrata una stima previsionale della durata totale della conversione in atto e di quella media per record; viene infine prodotto un *log file* col resoconto della procedura: tempi, record trattati, errori. In entrambi i casi i nomi dei campi non riconosciuti vengono mostrati a video proponendo di aggiungerli o trascurarli, di interrompere o continuare la procedura.

Data Magician organizza comunque ogni conversione di file in quattro distinte fasi:

- 1) definizione di tipo e struttura del file di input;
- 2) trattamento analitico dei campi in input;
- 3) definizione di tipo e struttura del file di output;
- 4) trattamento analitico dei campi di output.

Nella fase 1 c'è poco da dire, a parte il nome del file; quando si tratta di formati non aperti: (Marc Iso 2709 Star Inmagic dBase), fa quasi tutto il programma da sé; ci sono invece specificazioni cruciali da dare per i formati con delimitatori e ad etichette, a seconda dei casi: quanti sono i campi, cosa delimita i campi, cosa li separa, come si riconosce l'inizio e/o la fine di un record, come si riconoscono i nomi dei campi?¹³ La fase 2 è quella fondamentale: innanzitutto si vede se la prima è stata bene impostata e quindi se si riesce a leggere "qualcosa", prima i record, di- ➤

20. PROGRAMMA SOFTWARE: tipo di conversione fra formati <i>Input:Output n:n 1:1 1:n n:1</i>	<i>quasi</i> un convertitore puro n:n
21. PROGRAMMA: conversione preconfezionata o impostabile con specifiche e comandi di manipolazione	è impostabile, è preconfezionata per il formato del file (ad es. label e directory di MARC e ISO 2709 vengono letti e scritti automaticamente)
22. ISTRUZIONI CONDIZIONALI (SE...ALLORA) -può combinarle in espressioni -su più di un campo alla volta -può registrare i valori di verità verificati per richiamarli	sì; non ci sono parentesi, ma le condizioni si possono porre e verificare sequenzialmente "a cascata"; condizione rappresentata da una stringa presente/assente, da assenza/presenza di campo, dall'uguaglianza di un valore (alfabetico o anche numerico: numero di record, data...) dal confronto con un'altra condizione precedentemente posta in segnalini temporanei (fino a 99) e così via; condizioni ed elementi si legano con operatori booleani
23. PROGRAMMA: esecuzione in <i>batch</i> dalla linea dei comandi	sì
24. PROGRAMMA c'è controllo di correttezza sui comandi prima, durante, dopo l'esecuzione	sì, c'è controllo, e non solo sintattico, con messaggi di errore ed help, durante la scrittura, sia in input che in output, durante l'esecuzione, e dopo in un <i>log file</i>
25. PROGRAMMA: svolgimento della procedura può venire (a) interrotto; (b) visionato (1) complessivamente, (2) record per record, (3) con resoconto dell'esito	(a) sì, anche per verifica di condizioni; (b) sì, (1) complessivamente, (2) record per record con parziale interattività, (3) con resoconto finale in un <i>log file</i> ➤

stinti l'uno dall'altro, e poi i campi. Potendo, come appena detto, caricare un record alla volta, vedere come viene letto e poi come viene manipolato (e siamo ancora in input), ci si trova nell'osservatorio e laboratorio privilegiato. La fase 3 in genere non richiede alcun intervento, è una conseguenza della 1. La fase 4 può sembrare ridondante:¹⁴ può effettivamente ridursi anche ad un trascinare verso l'esito finale solo quello che ha preparato l'input, oppure può sovrapporre tutte le istruzioni che si vogliono, e svolgere un proprio lavoro. Fra l'altro, sia in input che in output si creano nuovi campi spostandovi il contenuto di uno o più di quelli già letti. Un'impostazione funzionale dei due gruppi di comandi in due fasi, è quella del macellaio, del norcino con cultura transregionale: tagliare, spezzettare, preparare i dati in input in modo che in output si possa ancora decidere di ricomporre diversamente in ragio-

ne di diversi assetti, perfino di diversi formati di destinazione: insaccati vari, carne fresca, carne da salare, congelare... così gli stessi dati vengono preparati in un solo modo e, copiando il file di filtro, possono venire elaborati in vari modi per output diversi; ovvero, a partire da uno stesso file e da uno stesso insieme di comandi di riformattazione nella fase 2 di input, si possono aggiungere comandi e specifiche di output diverse nella fase 4, per produrre file ad etichette, *comma delimited*, Iso 2709 ecc.

"The real power and flexibility of Data Magician is in the variety of processing options available".¹⁵ I comandi di riformattazione si applicano al contenuto dei file ovvero ai suoi record e dunque ai file di qualsiasi formato, una volta che questo è stato indicato e riconosciuto. Per documentare questi comandi, ed anche le funzioni generali del programma, e per fornire degli esempi, ho scelto di presen-

tare dati ed informazioni in quattro tabelle, che non sono integrative, ma sono la parte sostanziale di questa rassegna. La 1 è un sommario, incompleto e discorsivo, dei comandi disponibili; la 2 è una sorta di griglia valutativa che può arrivare a tracciare la carta d'identità di Data Magician;¹⁶ la 3 è l'esemplificazione commentata di alcuni comandi a fronte dei dati in input; la 4 è un esempio di trattamento di dati "non strutturati".

In definitiva, tutto bene, potente e semplice? Soliti angoli opposti al vertice: quanto più gli strumenti sono versatili, aperti alle combinazioni, tanto più richiedono che chi li usa voglia e sappia entrare nel vivo delle tecniche, per raccogliere in seguito frutti copiosi di un investimento non superficiale. Data Magician è senz'altro più difficile da usare dei programmi e delle funzioni di conversione incorporate in altri software per la gestione di archivi bibliografici: funziona bene,

Tab. 2 - Caratteristiche analitiche (segue)

<p>26. PROGRAMMA documentazione (a) manuale (b) help on-line (c) versione dimostrativa (d) tutorial per l'apprendimento</p>	<p>(a) manuale a stampa, a mo' di <i>reference</i>, con appendici per messaggi di errore ed indice analitico (b) help circostanziato (<i>context sensitive</i>), con indice e salto diretto alle varie pagine/schermate (c) sì, a pagamento (d) no, ma molti esempi del trattamento di interi file in formati diversi, su carta e su disco</p>
<p>27. PROGRAMMA: (a) facilità d'uso (b) potenza, flessibilità, precisione</p>	<p>(a) non è facile, né immediato, non richiede conoscenze di logica di programmazione; (b) le possiede tutte a buon livello</p>
<p>28. PROGRAMMA: (a) assistenza (b) manutenzione (c) requisiti per il funzionamento (d) costo (e) recapiti</p>	<p>(a) gratuita via e-mail dalla Folland Software Services; (b) non ci sono contratti; meno di una release all'anno; (c) IBM PC compatibile; 512k RAM; 350k RAM liberi; meno di 500k su disco per file di sistema + memoria di massa per file di input/output; stampante; (d) \$ US 25 demo (completamente funzionante, deducibili se si acquista la versione completa); \$ US 245 + spese postali (acquistando anche Library Master: 40% di sconto) (e) e-mail: lfolland@ca.dynix.com.; posta normale: 6 Chartwell Cresc. Guelph (Ontario) CA N1G2T-7</p>

fa cose in gran numero ed in varietà fantasiosa, quanto più lo si sa usare bene. Una documentazione più didascalica, capace di introdurre chi non ha esperienza di programmazione, e che colleghi le varie fasi, con la giusta enfasi sui preliminari cruciali e non con la sola illustrazione asettica tipica dei manuali di reference, non sarebbe di troppo. ■

Note

¹ Avvertenza: i nomi dei vari prodotti software citati sono marchi soggetti a copyright. Non intrattengo alcun rapporto commerciale o promozionale con la Folland Software Services. Manoscritto terminato il 18 luglio 1996; email: dellorso@unipg.it. Rimando al testo, ed ai riferimenti bibliografici ivi contenuti, di: *La conversione del formato dei dati: come, quando, perché. Riformattazione e downloading nella gestione degli archivi bibliografici*, "Biblioteche oggi", 12 (1994), 2, p. 24-31, a cura di chi scrive, a cui aggiungo da subito: F. DELL'ORSO - A. PANZA, *Téléchargement et reformatage de données Dobis/Libis pour la réutilisation dans un système local de gestion de fichiers*, p. 19, comunicazione presentata al 12th Annual Meeting of the Dobis/Libis Users Group, 14-17 September 1993, Coimbra, Portugal, con notizie su Data Magician, Fangorn, BiblioLink; per la pratica del lavoro (con EndNote per MacIntosh) cfr. S.N. BJØRNER, *Bibliography formatting software for managing search results: EndNote Plus and EndLink*, "Online", 1995, 5, p. 38-43; per un prototipo fatto in proprio: P.A. DEARNLEY, *Merging and reformatting records downloaded from various databases using database schemata mappings*, "Program", 29 (1995), 2, p. 159-168; nel volume miscelaneo curato da TERRY HANSON, *Bibliographic software and the electronic library*, Hertfordshire, University of Hertfordshire Press, 1995, cfr. E. CARTER, *A trial of pc bibliographic database and formatting packages*, p. 70-83, e, per i risvolti connessi al diritto d'autore: D. SLEE, *Electrocopying from databases*, p. 84-94, e C. OPPENHEIM, *Staying within the law*, p. 95-107.



² La Electronic Library Network in British Columbia, Canada, con 22 fra college e università ha creato grazie a Data Magician un database unico in formato Marc a partire da numerosi database nei seguenti formati: Marc (vari), Inmagic, dBase III, Ms Word; il Cut Catalogo unico toscano su cd-rom della Regione Toscana ha unificato in formato Iso 2709 dati provenienti da vari ambienti software; del pari sta accadendo ad alcune basi dati presenti in Internet sotto EasyWeb della Nexus di Firenze <<http://www.trident.nettuno.it>>.

³ Cfr. anche C.C. BROWN, *Creating automated bibliographies using Internet-accessible online library catalogs*, "Database", 17 (1994), 1, p. 67-71. Ho effettivamente percorso il ciclo virtuoso di ricerca, cattura, downloading via e-mail, riformattazione e caricamento agendo su Melvyl, Università della California, <telnet:melvyl.ucop.edu>.

⁴ Abituato anche a conversioni non perfette, resto ben più ottimista e soddisfatto di N. TANGARI, *Alcune implicazioni bibliografiche dello Standard generalized markup language (Sgml)*, "Bollettino Aib", 35 (1995), 4, p. 481-492, che parla di "traduzione, spesso complessa e di scarsi risultati", p. 484.

⁵ Cfr. S. STIGLEMAN, *Bibliography formatting software: an update buying guide for 1994*, "Database", 17 (1994), 6, p. 53-65 e M. ENGLE, *Downloading*

from the Melvyl® system to personal bibliographic databases, "Dla Bulletin", 12 (1992 Winter), 3, p. 18-20. Negli Usa è stato creato un Comitato Niso-National information standards organization con il fine di definire standard per il "Downloading from online systems"; i suoi lavori non sono conclusi.

⁶ Vedi S.N. BJØRNER, *Cleaning up search results: what do we expect?*, "Online", 1995, p. 65-68, e, per la persistente specifica superiorità di un editor sui word processor: S. UNGER, *Why patent searchers (and others) need Kedit when they already have a word processor, or, post-processing at the power level*, "Database", 17 (1994), 4, p. 63-7.

⁷ I primi due rispettivamente prodotti da Head Software International e da Tailored Information Ltd; Malachia è della Neotecnica di Verona (email: neotecvr@mbox.vol.it), versione 2.0 Windows, con help in linea in italiano; input, output e riformattazione sono parametrizzabili, atto a trasformare dati con formati discorsivi, tipo ISBD, in dati strutturati ad etichette, consentendo operatività interattiva o svolgimento batch. È in funzione presso l'Università di Siena.

⁸ IFLA, *ISBD(G)*, Roma, Aib, 1987, p. 11 0.1.

⁹ N.E. WILLE, *Retroconvertire con lo scanner: un'introduzione al progetto Facit*, "Bollettino Aib", 33 (1993), ►

4, p. 467-473. Dal momento che "Il software prototipale, insieme al suo manuale di accompagnamento, sarà di dominio pubblico per usi non commerciali..." e che "Il progetto, cominciato nel gennaio del 1993, durerà ventisei mesi..." è proprio vero che si sta avvicinando un momento quasi esaltante nella storia dei cataloghi a schede.

¹⁰ I suoi programmi di sviluppo prevedono ancora una versione Dos, la 1.5, l'aggiornamento della documentazione, e più oltre, forse, la riscrittura del programma sotto Windows95 in Visual Basic.

¹¹ Possibili miglioramenti di dettaglio nello scrivere le istruzioni di conversione: potere tenere sotto controllo quali segnalini sono stati finora utilizzati per registrarvi lo stato di condizione di "verità/falsità" e per vedere il valore di tale stato; potere cercare, e sostituire etc. stringhe di caratteri all'interno dei comandi stessi; potere duplicare rapidamente i comandi di input come comandi di output (anche

se è vero che in output i campi si possono richiamare anche solo per numero ed in input, si possono non indicare, affidandosi ad un riconoscimento automatico).

¹² R. PEREZ, *Software Leverage : moving documents information*, "Library Software Review", 12 (1993), 3, p. 10: "...superb trial-and-error inspection mode... This tool puts heavy-duty translation or conversion programming power in the hands of non-programmers".

¹³ L'approccio per il riconoscimento dei Tag è quello — funzionale ad una intensa attività di *downloading* da svariati database — di individuarli in base alla struttura indicata nelle specificazioni del file di input, in modo da non essere tenuti ad analizzare minuziosamente tutto il file, ma da tentare la intercettazione in base allo schema strutturale dichiarato. Così fa Library Master, che ha impostato la sua funzione di riformattazione su Data Magician, ma così non fanno Fangorn o EndLink. I comandi di output possono

poi leggere i campi riferendosi a quelli di input tramite il semplice numero di sequenza operativa. L'autore sta lavorando per aggiungere un diverso approccio ai Tag, tale per cui il programma prenda in considerazione solo quelli specificati da chi opera tramite il nome.

¹⁴ S. DETWILLER, *2,400 records and growing : using the Data Magician to add records to an Inmagic database*, "Online", 17 (1993), 1; p. 91: "The Output Fields screen is the most obscure part".

¹⁵ *The Data Magician version 1.4 user manual*, Guelph, Folland Software Services, 1992, p. 81.

¹⁶ *Check-list*, il cui punto di partenza, si parva licet, è stato il fondamentale contributo di E.G. SIEVERTS, *Software for conversion of downloaded data: criteria for comparison and assessment*, in *Online Information 89 : 13th International Online Information Meeting, London 12-14 december 1989: proceedings*, Oxford and New Jersey, Learned Information, 1989, p. 59-70.

Tab. 3 - Esempio di riformattazione dei dati

DATI IN INPUT E, SOTTO, I COMANDI DI RIFORMATTAZIONE	OUTPUT (tagged per Library Master)	COMMENTO
<p>Ti: Il Seicento // TG: Encyclomedia // = Le dix-septième // TP: ENCYCLO-MEDIA 1: Seicento // TG: [Enciclopedia multimediale su CD-ROM] // Opere didattiche</p> <p>Ti (vari titoli) BS" // "</p> <p>TIT: (titolo proprio) F"TI (1)" XS</p> <p>TSC: (titoli collettivi) F"TI (2-)" BX" TG:" S"TG:" XS s"[" s"]" "</p> <p>TPA: (titolo parallelo) F"TI (2-)" BX" = " S" = " XS</p> <p>TSP: (titolo particolare) F"TI (2-)" BX" TP:" S"TP:" XS</p> <p>TIOTHER (altro senza etichetta) F"TI (2-)" S"TG:" S" = " " S"TP:" * " XS</p> <p>Note: generale: Tit. orig.: C'était l'âge de Descartes collaz: 7 vol. edition: 3rd ed.</p> <p>Note</p> <p>NOTAG F"NOTE" C"generale: " BR"generale: " BL" " C~ XC</p> <p>COLLAZIO F"NOTE" C"collaz: " BR"collaz: " BL" " C~ XC</p> <p>EDIZIONE F"Note" C"edition: " BR"edition: " C~ XC</p> <p>Soggetti: Congressi - Firenze - 1996; Firenze - Storia - XVII sec.;</p> <p>Soggetti BS"; XS S" * - Firenze" FIRENZE --> * " XL</p> <p>Autore: LC - Finzi, Claudio</p> <p>AUTORE BS" - "</p> <p>A1 s1 f" autore" ST"c:\dm\lookup.tab"</p> <p>A2 s2 f" autore" ST"c:\dm\lookup.tab"</p> <p>Editore: Pergamon Press, Oxford -- Laterza, Roma</p> <p>Editore: Academic Press, New York</p> <p>Place: Toronto</p> <p>Pub: McGraw Hill</p> <p>Editore BS" -- " TB" ^n" s", " ^ " SM" " BM" ^ n ^"</p> <p>PLACE TB" ^ "</p> <p>PUB TB" ^n "</p> <p>IM: F"Editore" SF F"Place" F"Pub" XS</p>	<p>[T1:] Il Seicento</p> <p>[T2:] Encyclomedia</p> <p>Enciclopedia multimediale su CD-ROM</p> <p>[T3:] Le dix-septième</p> <p>[T4:] ENCYCLOMEDIA 1: Seicento</p> <p>[T5:] Opere didattiche</p> <p>[NG:] Tit. orig.: C'était l'âge de Descartes</p> <p>[CO:] 7 vol.</p> <p>[ED:] 3rd ed.</p> <p>[KW] FIRENZE --> Congressi - 1996 Firenze - Storia - XVII sec.</p> <p>A1: Cimmino, Luigi</p> <p>A2: Finzi, Claudio</p> <p>[IM:] ^ Oxford^nPergamon Press ^ Roma^nLaterza ^ New York^nAcademic Press ^ Toronto^nMcGraw Hill</p> <p>[TR:] Mono</p>	<p>frammentazione - parsing - dell'unico campo dei titoli; lo riduce in occorrenze; riconosce e separa le etichette, unisce quelle uguali (TG); sa trattare i titoli che non ne hanno alcuna</p> <p>frammentazione dell'unico campo delle note, cerca e riconosce inizio e fine di ciascuna; le archivia in campi distinti</p> <p>divide in occorrenze; estrae la stringa "Firenze", se interna, e la antepone in maiuscolo</p> <p>crea 2 occorrenze; le pone in campi distinti; confronta i dati con un file esterno, trovandovi "LC", prende il valore sostitutivo: "Cimmino, Luigi;"</p> <p>tratta 2 campi di nome uguale unendoli; tratta i sottocampi - apponendo identificatori, invertendo luogo ed editore, eliminando la punteggiatura -; tratta altri 2 campi fondendoli, ancora con codici di sottocampo; fonde tutto in un solo campo con 4 occorrenze e codici di sottocampo</p> <p>se c'è il campo TI è una monografia</p>

Tab. 4 - Riformattazione in formato Micro Cds/Isis (Iso 2709) di un file "non strutturato"

DATI IN INPUT

Davies, Ron -- Brandon, Ed
 Automated Thesaurus Management in a Network Environment
 In:
 Proceedings of the 22nd Annual Conference of the Canadian Association for
 Information Science
 Montreal: McGill, 1994, p. 326-332
 A description of some aspects of the OECD's MTM multilingual thesaurus
 management system
 \$
 Mattenella, L.E. -- Velazco, J.W.
 Diseño y desarrollo de una base de datos bibliograficos. <Design and
 development of a bibliographic database>
 Information Development, 8:233-6; Nov 1992
 CDS/ISIS application at the Institute for Minerals Development, Argentina
 \$
 Di Lauro, Anne -- Brandon, Ed
 Manual for Preparing Records in Microcomputer-Based Bibliographic
 Information Systems
 Ottawa: IDRC, 1990
 ISBN 0-88936-559-8
 \$

OUTPUT FINALE IN MICRO CDS/ISIS, TRAMITE ISO 2709

26: p
 16: ^aDavies^bRon
 16: ^aBrandon^bEd
 1: ^aAutomated Thesaurus Management in a Network Environment
 4: ^IMontreal^nMcGill^d1994
 5: 326-332
 10: Proceedings of the 22nd Annual Conference of the Canadian Association for Information Science
 7: A description of some aspects of the OECD's MTM multilingual thesaurus management system
 22: 1994

 26: a
 16: ^aMattenella^bL.E.
 16: ^aVelazco^bJ.W.
 1: ^aDiseño y desarrollo de una base de datos bibliograficos^pDesign and development of a bibliographic database
 10: Information Development^r8^dNov 1992^i233-6
 7: CDS/ISIS application at the Institute for Minerals Development, Argentina
 22: 1992

 26: m
 16: ^aDi Lauro^bAnne
 16: ^aBrandon^bEd
 1: ^aManual for Preparing Records in Microcomputer-Based Bibliographic Information Systems
 4: ^IOttawa^nIDRC^d1990
 8: 0-88936-559-8
 22: 1990

Sono arrivato ad elaborare questo file solo grazie all'impostazione data da L. Folland: ancora una volta la cosa fondamentale e più difficile era prendere un record e poterlo cominciare ad analizzare, dopodiché le analisi e le manipolazioni minute erano molto più facili. Per ragioni di spazio e di intellegibilità non riproduco i comandi che sono serviti alla riformattazione, ma il caso è reale e documentato. I capoversi nel file di input sono stati riprodotti come nell'originale ed hanno peso nellaformattazione. L'unico intervento manuale nel file di input è stato quello di aggiungere un segno di fine record "\$". Si notino i codici di sottocampo (campi 16, 1, 10, 4), l'aggiunta del campo 26, la duplicazione della data anche nel campo 22. Ho anche prodotto un output *comma delimited* ed uno ad etichette.