

Università degli Studi di Padova
Progetto di digitalizzazione del Polo Socio-economico

Relazione complessiva

Il **Progetto Digitalizzazione Riviste**, avviato dai Dipartimenti di Scienze Economiche “M. Fanno” e di Scienze Politiche “E. Ancheri” (in collaborazione con il Centro di Ateneo per le Biblioteche), si poneva inizialmente un duplice obiettivo:

- liberare scaffali occupati dai volumi cartacei
- dar vita a un esperimento di biblioteca digitale

Con queste finalità nel 2001 è stato stanziato un finanziamento per l'acquisto di attrezzature:

- strumentazione necessaria ad acquisire le immagini (scanner; workstation)
- server; software per l'archiviazione e la ricerca dei dati
- hardware per il recupero dei dati (postazioni di consultazione)

In seguito, grazie all'esperienza acquisita durante l'istruttoria della prima fase, il progetto è stato allargato a tutto il Polo Socio-Economico, con questi obiettivi:

- la digitalizzazione di una selezione di riviste italiane di particolare interesse storico, allo scopo di
 - 1) costruire un archivio accessibile in rete all'utenza, in formato elettronico (immagini e full-text) con la strumentazione già acquisita;
 - 2) verificare metodologie e costi di un più ampio progetto di digitalizzazione.

La scelta di uniformarsi a standard internazionali e a esperienze nazionali pregresse, appoggiandosi alla struttura di metadata di Dublin Core, assicura l'interoperabilità con altri digital repository a livello nazionale e internazionale.

Il progetto può avere come ricaduta secondaria lo sviluppo di archivi paralleli di materiale non bibliografico:

- 1) carte geografiche
- 2) dati statistici.

Il gruppo che ha lavorato all'avvio del progetto è partito da un'ipotesi di configurazione basata sull'acquisto e l'installazione di:

- 2 scanner piani (con relative workstation) per l'acquisizione
- 1 server potente per la gestione dei dati
- 1 web server, più 7 postazioni PC per la consultazione in rete.

La nostra attenzione si è alla fine concentrata su due ditte che si erano rese disponibili a rivedere insieme l'architettura del progetto. Con ciascuna di esse ci siamo incontrati in più occasioni per approfondire gli aspetti della loro offerta.

1. Acquisizione

La prima conclusione importante a cui siamo giunti è che, per digitalizzare le riviste in nostro possesso, lo scanner piano era uno strumento decisamente inadeguato.

Lo **scanner piano** fotografa il documento facendo scorrere un rullo sotto il piano di appoggio. Il documento, che nel caso di una collezione di riviste è quasi sempre un volume di notevole spessore, prima di ogni scansione deve quindi essere:

- aperto sulle pagine da scansionare
- posizionato a faccia in giù
- pressato con forza sul piano di appoggio, per ridurre al minimo l'effetto curvatura della pagina e le ombre che si producono
- centrato con cura

Queste operazioni manuali, ripetute per centinaia di volte su ogni volume, provocano non solo un'enorme perdita di tempo, ma anche il prevedibile sfaldamento della rilegatura del volume.

Per risolvere il problema occorre perciò rovesciare il sistema di ripresa, eseguendolo dall'alto.

Questo tipo di ripresa è propria di due strumenti di acquisizione: **il set di ripresa fotografica e lo scanner planetario.**

In entrambi i casi il volume viene posizionato una sola volta, a faccia in su, e centrato con cura rispetto al piano di ripresa. L'unico movimento a cui sottoporremo il volume sarà quello, quasi impercettibile, di sfogliarne le pagine prima di fotografarle: tutt'al più potrà essere necessario in qualche occasione pinzare i bordi delle pagine, per ottenere risultati migliori.

Abbiamo cercato di capire se il set di ripresa basato su una **fotocamera digitale** poteva essere adatto al nostro scopo, così ne abbiamo allestito uno avvalendoci di una fotocamera digitale messa a disposizione a Palazzo Maldura.

Il risultato è stato sorprendente per la qualità di immagine, positivo per la velocità di esecuzione, ma assolutamente carente per la correzione delle imperfezioni ottiche (curvatura della pagina, presenza di ombre, ecc). Ogni immagine prodotta avrebbe dovuto essere corretta manualmente intervenendo su centratura, contrasto, pulizia dei bordi, ecc.

Ci siamo così resi conto che per offrire uno standard di qualità paragonabile a quello delle biblioteche digitali presenti sul web era necessario affidarsi ad uno scanner planetario.

Lo **scanner planetario** in teoria possiede infatti tutte le funzioni fondamentali per produrre in tempi brevi un file d'immagine fedele all'originale, perfettamente centrato e allineato, perché presenta queste caratteristiche:

- ripresa dall'alto, condizione indispensabile per manipolare al minimo il volume
- presenza di un piano basculante che consente, una volta appoggiato il libro, di compensare il dislivello tra le due parti del volume in modo da metterle sempre in linea rispetto al piano di ripresa
- correzione ottica dell'immagine, necessaria a eliminare ombre, disallineamenti e curvature della pagina
- durata di vita: lo scanner planetario non ha né parti meccaniche usurabili (come il rullo dello scanner piano), né sensori deperibili (come il CCD della fotocamera digitale), per cui la sua vita media può essere stimata intorno ai vent'anni.

Nel mercato, a questo proposito, esiste uno scanner planetario in bianco e nero (formato A3, con possibilità di scansionare 2 pagine alla volta), che ha un costo compatibile con il finanziamento per questa prima fase del progetto: si tratta del [Minolta PS7000](http://www.minolta.it/digital/descrizione.asp?c=15) (<http://www.minolta.it/digital/descrizione.asp?c=15>).

Ci siamo recati presso la ditta [A. M. di Bologna](http://www.automicro.it) (<http://www.automicro.it>) per effettuare una prova sul campo del funzionamento dello scanner. Allo scopo avevamo portato un volume della rivista *La Riforma Sociale*, che a nostro parere assommava in sé tutte le casistiche che avrebbero potuto presentarsi durante il processo di digitalizzazione.

Lo scanner ha interessanti caratteristiche pratiche: soprattutto, oltre al piano basculante che riduce la curvatura dei libri, ha i pulsanti d'avvio della scannerizzazione collocati sul piano dove è posato il libro, quindi si può lavorare in piedi e con una certa velocità. A richiesta è dotato anche di un pedale opzionale.

Lo scanner è in grado, grazie a un sistema di specchi, di intervenire in maniera correttiva sulla curvatura del testo causata dalla rilegatura.

Lo scanner, tramite interfaccia TWAIN o ISIS, può interagire con tutti i programmi di visione immagini e di acquisizione presenti sul mercato.

L'abbiamo visto in azione con Photoshop e con Power Book (uno dei più importanti programmi di acquisizione).

2. Gestione

Il secondo aspetto importante per il progetto è quello relativo alla gestione e consultazione dei dati acquisiti con lo scanner.

Lo strumento necessario allo scopo è un software di archiviazione, attraverso il quale sia possibile applicare ai documenti digitalizzati quei campi di ricerca che permettano all'utente di recuperare l'articolo cercato.

Dai contatti avuti con le ditte coinvolte abbiamo potuto prendere visione di due proposte di software gestionale. Alla fine la nostra scelta è caduta su [DocManager](http://www.polysoft.it/doc_manager/) (http://www.polysoft.it/doc_manager/), perché presentava una struttura più aperta e maggiori potenzialità di sviluppo e di personalizzazione; una di queste è la possibilità di attivare l'OCR nella fase di indicizzazione delle immagini.

3. Consultazione

Il terzo aspetto riguarda la configurazione hardware necessaria a supportare il recupero dei dati. Nella ipotesi iniziale erano previsti 1 server per gestire i dati, 1 web server per distribuirli in rete e 7 postazioni per la consultazione, nonché una serie di licenze per i vari software necessari al trattamento dei dati acquisiti.

In realtà la decisione di puntare su uno scanner planetario e su un buon software di archiviazione ha ridotto notevolmente il budget destinato alla consultazione dati, che abbiamo preferito "congelare" per il momento, rimandando l'acquisto di server web e PC per la consultazione a una fase successiva del progetto.

La ripartizione del budget è quindi stata la seguente:

strumentazione per le fasi di lavorazione	Preventivo iniziale	Preventivo acquisto effettivo
Acquisizione (scanner, workstation)	22%	43%
Gestione (software di gestione, server)	24%	53%
Consultazione (web server, PC consultazione)	39%	3%
Licenze varie software	15%	1%

Informazioni sulla digitalizzazione raccolte durante la seconda fase istruttoria

Nel corso delle visite alle varie ditte che ci proponevano un'offerta, abbiamo acquisito un corposo dossier sulle varie tecnologie e attrezzature per la digitalizzazione, che può risultare utile ad altre strutture dell'Ateneo che intendano dedicarsi a tali progetti.

La ditta [SIAV](http://www.siav.it/Default.asp) (<http://www.siav.it/Default.asp>), con sede anche a Rubano (PD), si occupa di digitalizzazione testi per svariate strutture accademiche italiane. Trattando soprattutto materiale antico e/o pregiato, si avvalgono spesso di scanner molto più costosi di quello che acquisteremo noi, quali per esempio quelli della ditta [Zeutschel](http://www.zeutschel.de/) (<http://www.zeutschel.de/>), che produce anche attrezzature per il trattamento dei microfilm.

La ditta [A. M. \(Automazione Microfilm\)](http://www.automicro.it/) (<http://www.automicro.it/>) di Bologna offre sistemi di archiviazione su disco e microfilm e *service* di microfilm, scansione e plottaggio, vettorializzazione, masterizzazione; ha lavorato con tutte le principali università ed enti italiani.

Presso la A. M. abbiamo assistito a una dimostrazione del software [Book Restorer della i2S](http://www.i2s-bookscanner.com/en/products_software.asp) (http://www.i2s-bookscanner.com/en/products_software.asp), che interviene sulla consultazione, l'immagine e la pubblicazione e stampa dei libri digitalizzati. In particolare, con sofisticati algoritmi Book Restorer è in grado di correggere al 100% la curvatura presentata dalle pagine scansionate da libri rilegati. Questa funzione agisce anche in batch, quindi le operazioni di ripristino impostate si possono eseguire anche in assenza dell'operatore, durante le ore di chiusura.

Presso la sede della A. M. abbiamo inoltre raccolto numerosi [dépliant su scanner](http://www.automicro.it/sca_1.html) (http://www.automicro.it/sca_1.html) (tra cui i [DigiBook](http://www.i2s-bookscanner.com/en/products_microscanners.asp) a colori per libri antichi, http://www.i2s-bookscanner.com/en/products_microscanners.asp) e [programmi di digitalizzazione](http://www.automicro.it/image.html) (<http://www.automicro.it/image.html>).

Durante l'acquisizione con lo scanner si possono trattare le immagini con vari software, tra cui i più noti e diffusi (es. Imaging, Photoshop, PixUtil, Irfanview) e altri, come PowerBook e [AMBook](http://www.minolta.it/digital/accessori.asp?i=246&c=15) (proposto dalla A. M. in tre versioni <http://www.minolta.it/digital/accessori.asp?i=246&c=15>).

Anche in questo caso, in un primo tempo ci avvarremo dei software a nostra disposizione, in seguito ci riserviamo di ricorrere a un software di questo tipo, ma solo se si rivelerà strettamente necessario.

Attività già svolte:

- ✓ Due mesi di istruttoria sui macchinari con missioni dai fornitori e incontri in sede
- ✓ Vari mesi di studio dei progetti di digitalizzazione europei ed italiani ed adozione degli standard "open archive"; implementazione dei risultati alla nostra configurazione di archiviazione, in collaborazione con la ditta Polysoft
- ✓ Definizione della tipologia e del formato dei documenti da inserire

- ✓ Definizione del programma da utilizzare per l'acquisizione delle scansioni (scelta tra Irfanview, PixUtil, Imaging, Photoshop dopo una sperimentazione)
- ✓ Controllo dei descrittori associati ai dati immessi nel programma di archiviazione
- ✓ Definizione della procedura di inserimento documenti
- ✓ Personalizzazione dell'interfaccia Web utente per la ricerca e il *browsing*
- ✓ Stesura dei manuali e delle guide per la digitalizzazione ad uso del gruppo di lavoro
- ✓ Sperimentazione delle opzioni di digitalizzazione (scanner e programmi di trattamento immagini)

Seconda fase:

Con gli stanziamenti per i progetti speciali del CAB, uniti a un cofinanziamento di vari dipartimenti del Polo, si sta procedendo all'acquisto di attrezzature e risorse complementari a quelle già presenti per il progetto (software, scanner piano a colori, PC dedicati alla consultazione, stampante laser B/N di alta qualità, licenze software, contratti di assistenza, accessori, eventuale personale a contratto per il progetto). In seguito, dopo la fase pilota (che si sta rivelando più lunga del previsto), il progetto verrà esteso alle altre strutture del Polo.

Si è evidenziata subito **una voce di spesa che non avevamo previsto**: l'esborso per macchinari e programmi di alta qualità non si esaurisce al momento dell'acquisto, ma diventa continuativo a causa dei contratti di assistenza e manutenzione, degli aggiornamenti del software e di eventuali accessori da aggiungere volta per volta. Chiunque intraprenda un progetto del genere, quindi, dovrà tenere conto di questo aspetto e prevedere un flusso di finanziamenti continuo nel tempo.

Flusso di lavoro.

Abbiamo innanzitutto indicato, a livello del nostro Dipartimento e del Polo Socio-Economico, quale e quanto sarà il personale destinato a lavorare al progetto, in un team di bibliotecari e informatici.

È in corso ora un periodo di prova (a partire dalla messa a regime dei macchinari e dalla fine della fase di addestramento all'uso del software) in cui abbiamo scelto vari campioni di documenti da scansionare.

Il materiale scelto dovrà ovviamente essere libero da problemi di diritto d'autore. A tale scopo è necessario scegliere riviste con articoli i cui autori siano deceduti da più di 70 anni, e comunque riprodurre in rete solo le immagini delle stesse, in formato non modificabile (i più consigliati sono

TIFF e JPG) e senza alterare la struttura originaria della rivista (di qui la necessità di produrre dei file multipagina per ciascun articolo).

Abbiamo sperimentato i vari software grafici per l'acquisizione e il trattamento di immagini:

Problemi di PixUtil: non ha la funzione di rifilatura; l'interfaccia con lo scanner è pessima; la visualizzazione dell'immagine a schermo è meno pratica rispetto a Imaging

Problemi di Irfanview: la visualizzazione dell'immagine a schermo è meno pratica rispetto a Imaging, e soprattutto di qualità pessima, quindi non ci si può rendere conto sul momento della qualità dell'immagine prodotta. Il rischio è dover rifare tutto in seguito. Ha però un'utile funzione di rifilatura in automatico, che può tornare utile in futuro.

Problemi di Photoshop: non supporta i TIFF G4 e i multipagina

Alla luce di tutto questo si è scelto di utilizzare il programma Imaging della Kodak, disponibile nella dotazione standard di Windows 2000 ma non più in quella di Windows XP.

Una volta inquadrati tutti i problemi e stabilito un flusso di lavoro costante con una metodica consolidata, si potranno stabilire i tempi e i modi per il completamento del progetto, per un'eventuale acquisizione di altri software per il trattamento delle immagini e per l'allargamento del parco macchine necessario alla consultazione.

Problemi:

1) In osservanza a quanto stabilito dalle norme vigenti sul **diritto d'autore**, inizialmente sarà possibile mettere a disposizione le riviste solo all'interno del nostro Ateneo.

2) Si rende necessario **trovare adeguata collocazione definitiva alle attrezzature** per il progetto. Appena possibile si prevede di allestire dei locali dedicati al progetto nella sede di Ca' Borin, quando saranno ultimati i lavori di restauro della stessa.

3) **Lo scanner planetario**, contrariamente a quanto garantito dal fornitore, **non è in grado di trattare con precisione volumi di formato anomalo** (inferiori al B5, per esempio) e con una rilegatura molto alta e stretta (come quasi tutte le riviste che abbiamo scelto). Abbiamo quindi dovuto sperimentare vari artifici per facilitarne la scansione. Purtroppo con una lastra di plexiglass non si ottengono benefici, perché non è abbastanza pesante da permettere a chi scansiona di effettuare l'operazione senza dover tenere premuto il libro; si dovrà scartare la lastra di vetro per motivi di sicurezza. Abbiamo acquistato il pedale per lo scanner offerto dalla ditta A. M., ma non risolve comunque i problemi di errori. Sono in corso trattative e istruttorie per risolvere queste difficoltà tecniche.

aggiornato al 21 ottobre 2003

a cura della dott. Anna Mioni (collaborazione di Stefano Varotto per la fase istruttoria)

Biblioteca, Dipartimento di Scienze Economiche "Marco Fanno", Università di Padova
e-mail: anna.mioni@unipd.it sito internet <http://cordialmente.cab.unipd.it:8080/CAB/progetti/5>