

# Il Münchner Digitalisierungszentrum e lo stato dell'arte degli scanner automatici

Commissione digitalizzazione BESS<sup>1</sup>

[info@bess-piemonte.it](mailto:info@bess-piemonte.it)

*In una tre giorni organizzata dalla Bayerische Staatsbibliothek  
presentate alcune interessanti novità in questo settore*

Klaus Kempf<sup>2</sup>

[klaus.kempf@bsb-muenchen.de](mailto:klaus.kempf@bsb-muenchen.de)

Ventottomila titoli full-text on-line e più di 80 progetti di digitalizzazione completati. In queste due cifre sono evidenziate le dimensioni del Centro di digitalizzazione (MDZ, Münchner Digitalisierungszentrum) della Bayerische Staatsbibliothek di Monaco (BSB),<sup>3</sup> ma MDZ è ben più di questo.

L'offerta di documenti on-line di MDZ è partita nel 1997 grazie ad un programma di sostegno della Deutsche Forschungsgemeinschaft (DFG) che mirava allo sviluppo della ricerca scientifica e delle competenze tecniche delle biblioteche tedesche in questo ambito. Oggi in Germania numerose biblioteche si occupano di digitalizzazione, ma i due centri di digitalizzazione più importanti rimangono MDZ e il Centro di digitalizzazione dell'Università di Goettingen (GDZ) che è stato istituito anche nel 1997 come secondo centro nazionale di competenze sulla digitalizzazione.

L'attività di produzione di contenuti digitali on-line che la BSB conduce si colloca nel contesto di una grande biblioteca universale di ricerca che agisce a livello internazionale con un compito speciale regionale di formazione. Per

la ricchezza e la qualità del suo patrimonio (9 milioni di volumi, di cui circa 110.000 manoscritti e incunaboli e una crescita annuale di circa 150.000 volumi) la BSB costituisce la seconda biblioteca della Germania occupando circa 700 persone che assicurano un servizio a 360 gradi. La sua collezione è particolarmente rivolta alle scienze umanistiche e sociali, possiede inoltre un ricco fondo di periodici nell'ambito delle "life sciences" e numerosi fondi speciali tra cui carte geografiche antiche, un archivio fotografico e una ricca collezione di fondi musicali. La biblioteca, fondata 450 anni fa, è oggi aperta sette giorni su sette dalle 8 alle 24 e offre 750 posti di lettura. Il prestito (1,8 milioni l'anno) è automatizzato e la BSB, grazie al suo straordinario fondo di periodici ricco di più di 40.000 titoli correnti, è il più grande fornitore di SUBITO, il servizio tedesco di prestito interbibliotecario e document delivery, con circa 350.000 pezzi all'anno. Nell'ambito dei compiti che le sono affidati a livello federale la BSB gestisce diverse biblioteche specialistiche virtuali, fra cui quelle di musicologia, storia, cultura e storia dell'Europa Orientale. Infine la BSB

svolge il compito di biblioteca di deposito per la Baviera e pubblica la bibliografia regionale bavarese, con un incremento annuo di circa 25.000 citazioni.

## Il Münchner Digitalisierungszentrum (MDZ)

La dotazione di macchine del centro di digitalizzazione è proporzionata alle dimensioni dell'istituzione da cui dipende e al compito che questa si è data (cfr. Tab. 1). Il laboratorio di riproduzione dispone di 15 scanner e macchine fotografiche dedicate per le riprese di originali che necessitano di particolari tecnologie e precauzioni di manipolazione. La definizione degli impianti arriva a 600 dpi e le dimensioni degli originali riproducibili fino ad A0. La gamma degli impieghi varia dalla riproduzione ad alta definizione di carte geografiche ai manoscritti di grandi dimensioni e con particolari esigenze di conservazione agli scanner dedicati alla riproduzione veloce di originali dalle caratteristiche differenziate (dai manoscritti agli originali a stampa). Oltre a questa dotazione speciale, MDZ

dispone di numerosi scanner planetari e convenzionali e macchine fotografiche digitali in grado di soddisfare la maggior parte delle esigenze poste da una collezione così ampia e differenziata come quella della Biblioteca di Stato del-

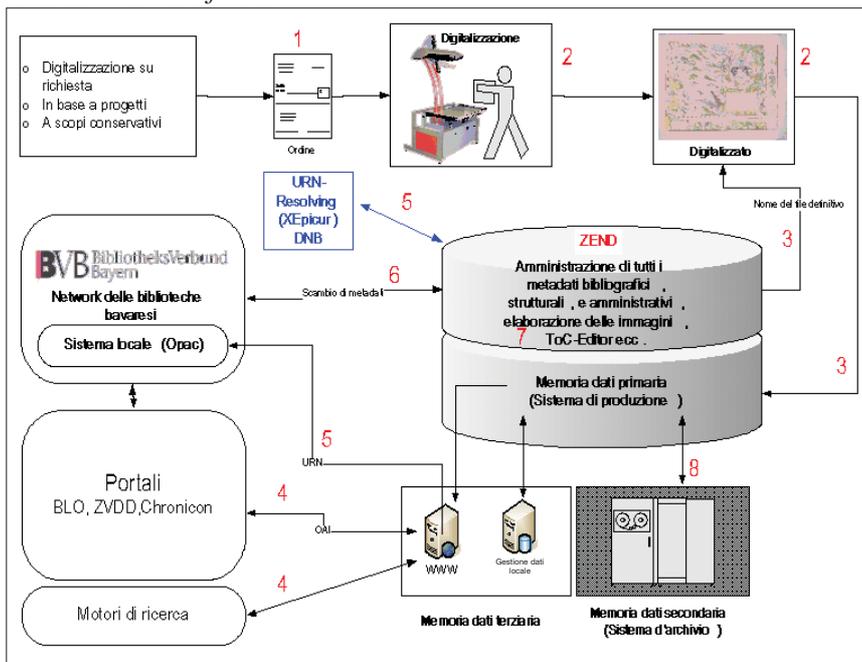
la Baviera. La dotazione di personale è tuttavia appena sufficiente alle dimensioni di MDZ. Sotto questo profilo il centro di Monaco deve appoggiarsi in larga misura a personale "a progetto", mentre il personale a tempo indeterminato

rappresenta una quota minoritaria. Sin dall'inizio MDZ ha lavorato nella convinzione che la digitalizzazione su larga scala richieda un sistema completamente automatico di gestione del *workflow*. Bisogna cioè che sia possibile controllare ogni singola fase individuale del processo di lavoro con totale trasparenza per l'intera struttura. In tal modo è possibile a chiunque sapere con precisione in ogni momento in che stadio di lavorazione si trovi un particolare documento. A questo scopo dal 2004 MDZ usa in modo sistematico uno strumento di gestione dell'intero flusso di lavoro e di amministrazione della banca dati informativa delle attività in corso (ZEND: Zentrale Erfassungs- und Nachweisdatenbank, cfr. Fig. 1). ZEND coordina in modo automatico i processi di pre-produzione, produzione e post-produzione attraverso i quali i documenti devono passare: gestione del lavoro, conversione del formato dell'immagine, catalogazione, chiusura del lavoro e archiviazione a lungo termine. I componenti dell'applicativo ZEND sono stati realizzati impiegando software *open source* e progettati per poter girare indipendentemente dall'architettura hardware e software della piattaforma. L'utilizzo di un'interfaccia XML consente il controllo delle varie fasi dell'iter del documento digitale attraverso varie funzioni di ricerca così come l'ordinamento logico del materiale e la sua collocazione nei vari servizi web, dall'OPAC fino ai diversi portali tematici in "costruzione".

Tab. 1 – Dotazioni del Münchner Digitalisierungszentrum

Quantità	Tipo di macchina	Formato max.	Definizione (dpi)
2	Scanner tipo "Grazer Kameratisch" per il trattamento degli originali più rari e preziosi	A3	600
3	ScanRobot SR300 Treventus. Due di queste macchine vengono utilizzate per il progetto "Stampa del 16. sec."	A3	300
1	Fotocamera Linhof completa di stativo e illuminazione. Viene utilizzata per oggetti tridimensionali negativi e altro materiale trasparente	A3	400
1	Stativo con fotocamera digitale portatile per lavori fuori sede. Modello TCCS 4232 prodotto dal centro VESTIGIA dell'Università di Graz	Fino a 42x32 cm	A seconda del modello di fotocamera
1	Scanner per microfilm Microbox della Book2net	A2	400
1	Scanner per carte geografiche e grandi formati	A0	400
6	Scanner planetari Zeutschel	A1	600
Vari	Scanner flatbed e fotocamere digitali	A3	2400

Fig. 1 - Schema del flusso di lavoro del Münchner Digitalisierungszentrum coordinato dal software ZEND



## Gli scanrobot

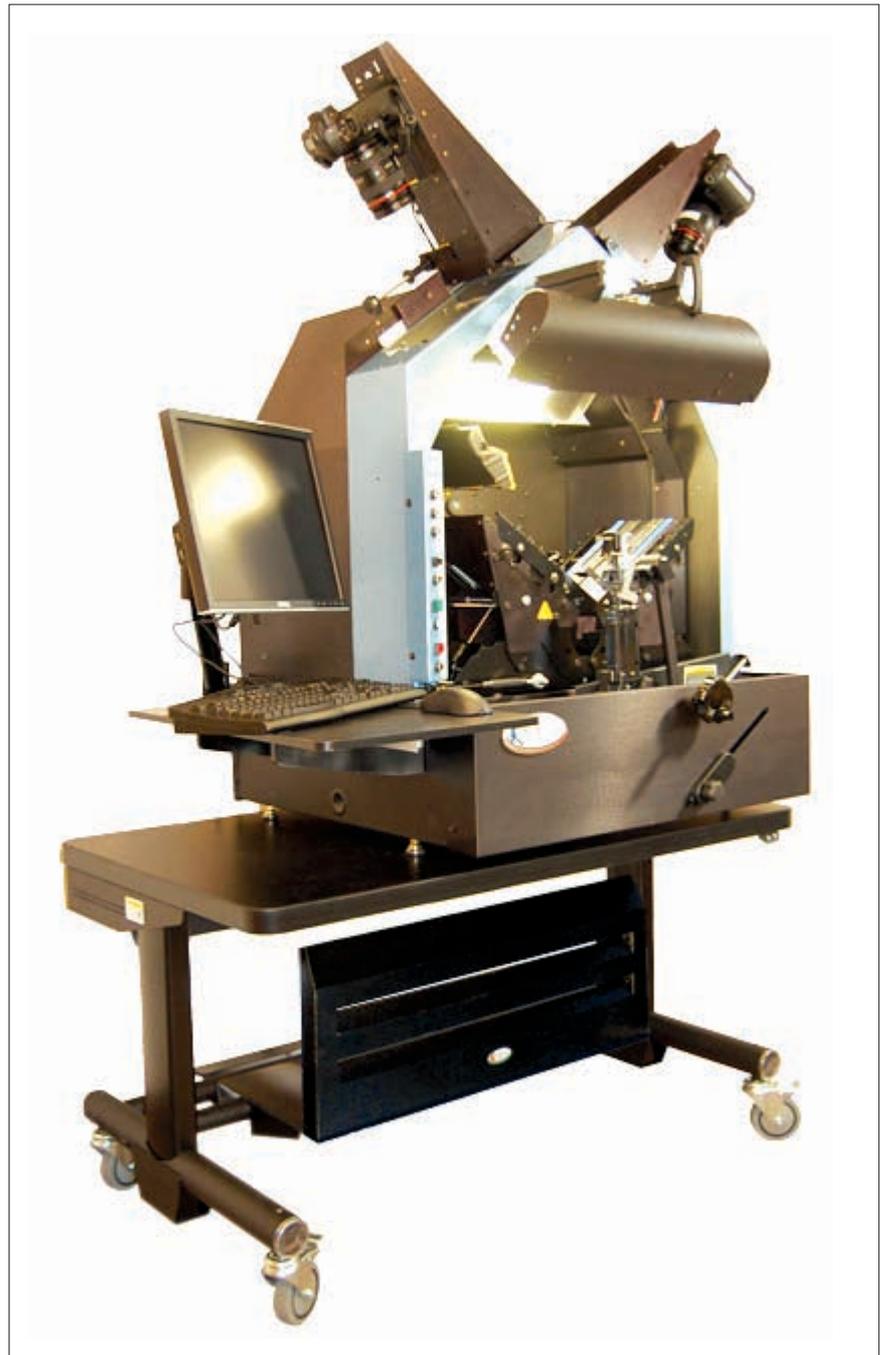
Nel quadro dei propri compiti di ricerca scientifica e sviluppo delle tecnologie della digitalizzazione MDZ (insieme all'Istituto per il restauro del libro e del manoscritto della BSB) ha organizzato nei gior-

ni 18-20 di giugno, presso la sua sede, una tre giorni dedicata alla presentazione e al confronto dei prodotti di quattro aziende specializzate nello sviluppo di scanrobot e del relativo software di digitalizzazione (cfr. Tab. 2). Nel corso della presentazione è stato possibile osservare dal vero il funzionamento delle macchine ed interrogare i tecnici sulle specifiche operative. Alla presentazione hanno assistito le biblioteche dell'Ires e Gioele Solari della Facoltà di Scienze Politiche dell'Università di Torino, che fanno parte della commissione digitalizzazione del gruppo di cooperazione di BESS.

La tecnologia applicata in questo settore deriva dalla convergenza di componenti in parte sviluppati a scopo diverso e in parte *ad hoc*; inoltre, l'elevata specializzazione delle macchine non impedisce che le caratteristiche e le funzionalità dei progetti riflettano per ognuno dei modelli una specifica destinazione. L'origine e lo sviluppo delle imprese costituisce, tra l'altro, un interessante *case study* di innovazione e di elevata specializzazione creativa.

Le imprese sono: 4DigitalBooks – ASSY SA;<sup>4</sup> Kirtas Technologies, Inc;<sup>5</sup> Qidenus Technologies GmbH<sup>6</sup> e Treventus Mechatronics GmbH.<sup>7</sup>

La prima azienda è frutto dello sviluppo dell'esperienza nel settore della riproduzione digitale della ditta francese i2S Innovative Imaging Solutions. La tecnologia di ripresa utilizza ottiche senza otturatore sviluppate dall'azienda francese che, essendo prive di parti in movimento, presentano il vantaggio di assicurare maggiore velocità e minore usura. Kirtas è uno *spin-off* della Xerox. Qidenus nasce come azienda specializzata nell'offerta di bracci automatici per sfogliare libri e spartiti per disabili e musicisti. Infine Treventus è una *start-up* dell'Università Tecnica di Vienna e si è valsa della stretta collaborazione



**Vista complessiva dell'APT BookScan 2400 della Kirtas**

con la BSB (ancora in corso) per la messa a punto della macchina.<sup>8</sup>

#### *Il principio di funzionamento*

Gli scanrobot possono essere classificati in vario modo, ma per semplicità espositiva qui vengono raggruppati in base alla tecnica per la manipolazione della pagina e al sistema di ripresa. Il sistema di ri-

presa può adottare indifferentemente un'ottica fotografica o una testa di scanner. La soluzione fotografica è quella prescelta da Kirtas e Qidenus, quest'ultima, in particolare, offre la scelta tra una fotocamera consumer o una macchina industriale ad alta definizione. In questo caso una o due macchine fotografiche digitali sono montate

Tab. 2 - Caratteristiche di base dei modelli presentati alle Scan Roboter Tage (Monaco, 18-20 giugno 2008)

	<b>4 DigitalBooks DL 3000</b>	<b>Kirtas APT 2400</b>	<b>Qidenus RBS Pro</b>	<b>Treventus SR 300</b>
Prezzo indicativo senza IVA in € (per tutte le macchine è anche possibile l'affitto e il leasing)	190.000 B/N 210.000 Colore	110.000	125.000	90.000
Dimensioni (cm) Lun. Larg. Alt.	301 x 150 x 220	84 x 76 x 122 (senza tavolo)	88 x 95 x 180	80 x 80 x 192
Peso	1 tonn.	75 kg.	200 kg.	180 kg.
Alimentazione	Trifase 400 V. 10kVA	Alternata 120/240 V. x 15A	Alternata 220 V.	Alternata 220 V. x 8A
Dimensioni pagina	Da 105x40 mm a 420x594 mm	Da 115x178 mm a 280x 355 mm (330 dpi) Da 115x178 mm a 230x307 mm (400 dpi)	Da 100x150 mm a 280x350 mm	Da 100x100 mm A 300x320 mm
Dimensioni dorso max.	150 mm	101 mm	100 mm	120 mm
Apertura libro	180 gradi	110 gradi	60, 90 e 120 gradi	da 60 a 120 gradi
Grammatura max	Da 20 g/m2 a 300	Da 49 g/m2 a 300	Non ci sono limiti purché la pagina non sia rigida	Da 60 g/m2 a 135
Velocità	Da 1,2 sec. a 4 per pagina	2.400 immagini/ora (teoriche)	2500 immagini/ora (teoriche) 1.500 realistico	valori pratici: 1.500 pagine/ora media e massimo di 2.400 pagine/ora a seconda dimensioni libro
Ottica	Tecnologia i2S CopiBook 56 mp per ogni obiettivo	Canon EOS -1Ds Mark III 21.1 mp	Canon EOS DSLR Fotocamere industriali ad alta definizione	Treventus PCD300™ (Treventus prism and capturing device)
Risoluzione	Da 200 a 600 dpi	Da 330 a 400 dpi. Il 600 dpi è possibile per libri di piccole dimensioni	Da 300 a 600 dpi a seconda della fotocamera	300 dpi (indipendente dal formato del libro)
Profondità colore	16 M. colori	-	-	30 bit
Software	Abby Recognition Server	APT Manager e Bookscan Editor; è possibile acquistare separatamente barcode scanner per MARC, Dublin Core, ecc. e scaricare metadata	QiScan Workflowmanagement + Abby and Imageprocessing routines	ScanGate™ +Workflowmanagement + Batch- e Postprocessing + XML Metadata (METS) opzionale: - OCR di Omnipage e ABBY - barcode scanner
Referenze	U.Stanford /USA U. Southampton /UK Infotechnique /FR Jouve /FR Nat.Lib. Slovenia Nat.Lib. Finland Nat.Lib. Norway Nat.Lib. Slovakia	400 clienti in 20 paesi, tra cui: Università di Bologna, National University of Ireland, Yale University, e United Nations	Medical University Vienna Business Administration Univ. Vienna	MDZ Monaco GDZ Gottinga Biblioteca Universitaria del Tirolo e Graz
Note dei produttori		Offre un programma per ristampe e la vendita di ristampe attraverso canali come Amazon		Il software "ScanGate™ light", può essere acquistato separatamente, e consente il pre- e postprocessing di ScanJobs indipendente- mente dallo ScanRobot SR300

superiormente alla macchina e posizionate in modo che la culla di sostegno del libro presenti la pagina in modo normale rispetto al campo di ripresa (nel caso della

macchina più economica della Kirtas che è dotata di una sola fotocamera, le pagine vengono presentate al piano di ripresa tramite uno specchio basculante). Nel ca-

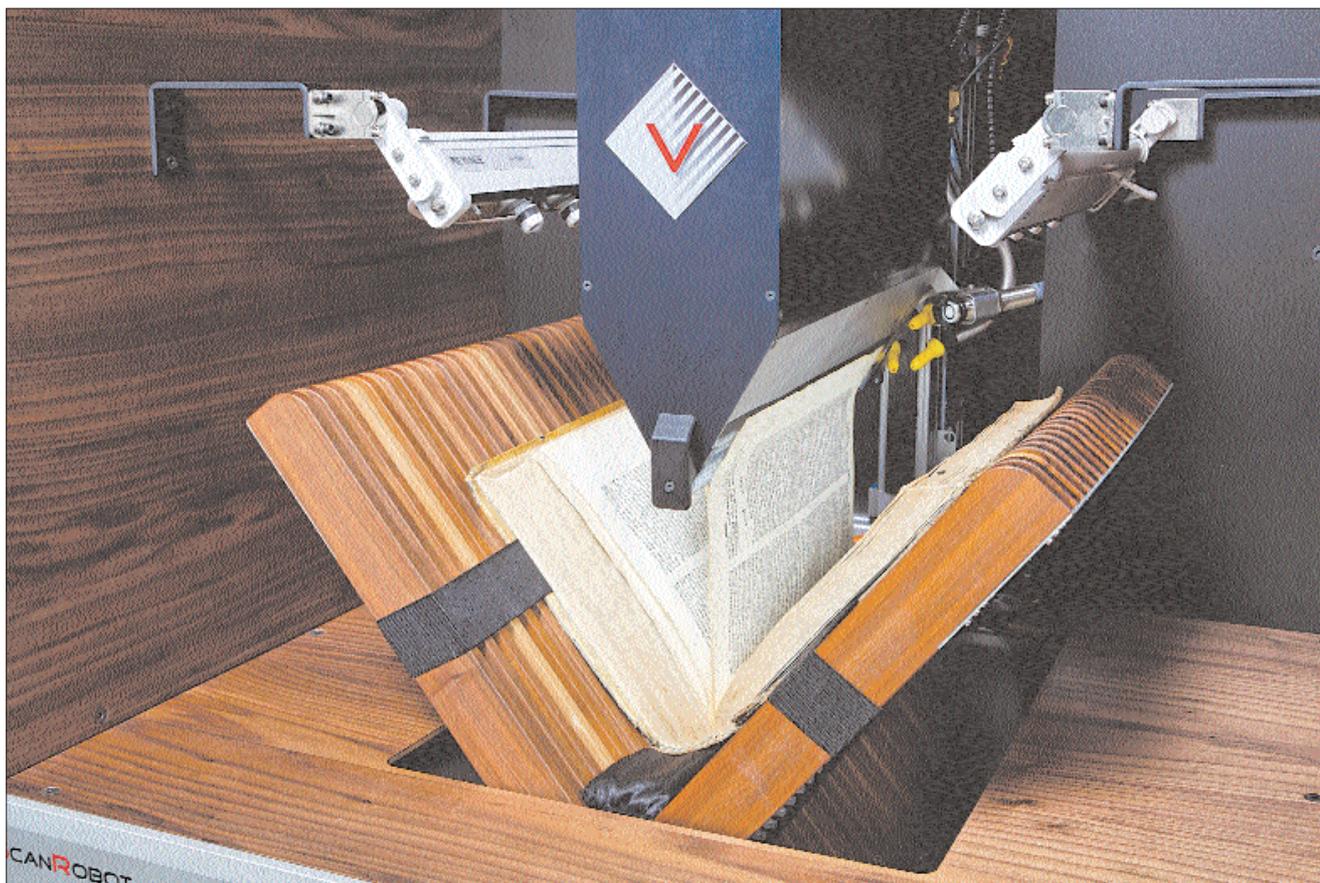
so del sistema proposto dalla 4DigitalBooks la testa ottica di ripresa è collegata all'hardware di elaborazione senza otturatore. Infine, il modello proposto da Tre-

ventus è l'unico che utilizzi una testa di scanner che scorre lungo le pagine e le riprende contemporaneamente tramite un prisma interposto tra di esse. Usando questo prisma, è possibile riprendere le immagini senza distorsione. Così non è necessaria alcuna correzione via software. Inevitabilmente le soluzioni adottate presentano vantaggi e svantaggi. Le fotocamere digitali vengono normalmente accreditate con garanzia fino a trecentomila cicli, dopodiché la parte meccanica dell'otturatore non offre più le garanzie di efficienza richieste e può rendersi necessaria la sostituzione del dorso. In questo caso è però evidente che si tratta di componenti per il mercato "consumer" con prezzi tendenzialmente in discesa. D'altro canto anche gli altri sistemi presentano curve di obsolescenza, sebbene meno ripide. Un vantaggio offerto dalla soluzione 4Digital Books è la possibilità di arrivare a 400 dpi senza interpolazione. La soluzione Treventus, essendo basata su scanner, offre invece il vantaggio di assicurare i 300 dpi indipendentemente dal formato; inoltre, per lo stesso motivo, non richiede illuminazione esterna.

L'aspetto ovviamente più innovativo e quindi più interessante è il meccanismo predisposto per sfogliare le pagine. Qui ci troviamo di fronte a soluzioni proprietarie, anche se i principi sono fondamentalmente due: *vacuum* e/o flussi d'aria per 4DigitalBooks, Kirtas e Treventus; meccanico per Qidenus. Inoltre è necessario specificare che è diversa la soluzione anche per la culla, cioè il meccanismo di sostegno del libro: salvo 4DigitalBooks che presenta l'originale aperto a 180° risolvendo in post-produzione la curvatura della pagina in corrispondenza della legatura, gli altri scanrobot optano per una culla ad apertura inferiore: 110° per Kirtas, 60°, 90° o 120° per Qidenus e da

60° a 120° per Treventus. Quest'ultimo approccio offre il vantaggio di non danneggiare la legatura dei libri e non richiede elaborazioni per la correzione di eventuali curvature della pagina. Non è semplice far comprendere come in realtà operano i differenti meccanismi e per questo è indispensabile poter osservare gli scanrobot all'opera, come è stato possibile fare a Monaco. Tuttavia si descrive in modo estremamente sommario la sequenza per dare un'idea di massima del funzionamento.

Il sistema adottato da 4DigitalBooks si basa sull'uso contemporaneo di un meccanismo per la creazione del vuoto, di un braccio meccanico e di un soffio d'aria. In buona sostanza la pagina viene aspirata da un vuoto pneumatico che la solleva. Successivamente viene sostenuta da un soffio d'aria che la mantiene in posizione tra due sbarre finché non viene completamente girata grazie ad un rapido movimento di traslazione orizzontale. In questo caso la velocità dell'operazione può essere adattata in base alle caratteristiche fisiche della carta. Kirtas impiega un braccio che crea un vuoto pneumatico per sollevare la pagina la quale viene successivamente girata dal movimento del braccio stesso. Treventus usa un doppio principio che consente la ripresa in contemporanea delle immagini di due pagine: un doppio flusso d'aria crea la necessaria pressione affinché le due pagine dove è aperto il libro aderiscano alla testa prismatica dello scanner il quale, spostandosi dal basso verso l'alto, genera in successive sequenze le immagini. Al termine della ripresa, il flusso d'aria viene diretto in modo tale da girare le pagine e riavviare il ciclo. Qidenus è invece l'unico scanrobot ad affidarsi ad una tecnologia completamente diversa. In questo caso si tratta del cosiddetto "dito bionico" che, esercitando una



**Treventus SR300: vista della culla di ripresa e del prisma ottico**

pressione microregolata, agisce come un vero dito umano. Il meccanismo, dotato di una testa in gomma intercambiabile a seconda delle caratteristiche della carta, applica l'opportuna pressione sull'angolo inferiore destro della pagina e la fa scorrere nella posizione desiderata.

Tutte le macchine sono dotate di particolari accorgimenti per evitare che vengano sollevate più di una pagina per volta o che l'operazione fallisca mancando di prelevare la pagina successiva. Quando tale evento si produce la macchina si ferma e richiede l'intervento dell'operatore. In alcuni casi sono previste operazioni semiautomatiche per segnalare fogli mobili o pagine multiple che necessitano un trattamento *ad hoc*.

### *Il software*

In base alle dichiarazioni dei pro-

ductori, tutti gli scanrobot sono dotati di un software proprietario per gestire il settaggio delle operazioni di ripresa (dimensioni della pagina, spessore del volume e della carta, contrasto necessario ecc.) e per svolgere l'attività di post-produzione (pulitura dell'immagine, scontornamento dei bordi, raddrizzamento della pagina, creazione dei formati di presentazione e OCR).<sup>9</sup> Il software di settaggio è particolarmente importante perché, nel caso di caricamento di volumi con caratteristiche assai diverse, può rallentare significativamente le operazioni. In tutti i casi è necessaria un'imputazione manuale di alcuni dei parametri che definiscono il profilo dell'originale. I parametri possono essere memorizzati in un database e all'occorrenza richiamati. Tutti i produttori forniscono un dato indicativo per quanto riguarda i tempi necessari per questa operazione: 4DigitalBooks e

Treventus segnalano che la maggior parte dei parametri viene rilevata automaticamente dalla macchina e quindi i tempi morti tra un libro e l'altro sono ridotti al minimo. Kirtas indica un tempo indicativo tra i 3 e i 5 minuti. Bisogna comunque considerare che, nel caso di grandi serie di volumi con caratteristiche similari questi tempi si possono ridurre significativamente.

Il software per la post-produzione riveste altrettanta importanza di quello per la gestione della riproduzione. In alcuni casi tale software viene venduto separatamente e può essere proficuamente utilizzato con altri, meno costosi, sistemi di ripresa. È opportuno precisare che di norma esso si limita alle operazioni di pulitura della pagina e alla creazione dei formati desiderati e include solo opzionalmente la gestione dei metadati e la presentazione della pagina online.

Questi aspetti infatti dipendono in larga misura dalle caratteristiche del progetto e dalle scelte degli standard operate dalla biblioteca o dal cliente (METS/ALTO, MAG ecc.).

#### Note

<sup>1</sup> Giorgio Bertolla, Tommaso Garosci, Marina Grazioli, Domenico Parigini. La commissione digitalizzazione di BESS (Biblioteca elettronica di scienze sociali ed economiche del Piemonte) fruisce del sostegno della Compagnia di San Paolo di Torino che da tempo ne finanzia le attività relativamente a diversi progetti.

<sup>2</sup> Bayerische Staatsbibliothek, Monaco di Baviera.

<sup>3</sup> <www.muenchener-digitalisierungszentrum.de>.

<sup>4</sup> Champs Courbes 23, 1024 Ecublens, Svizzera, <www.4digitalbooks.com>. Contatto: Ivo Iossiger, ivo.iossiger@4digitalbooks.com

<sup>5</sup> 7620 Omnitech Place, Victor, N.Y., 14564 USA, <www.kirtaseurope.com>.

Contatto: Joop Geesinkweg 999, 1096 Amsterdam, Marcel Aldewereld, mal-dewereld@kirtas.com

<sup>6</sup> Nedergasse 25/5, 1190 Wien, Austria, <www.qidenus.com>. Contatto: Vitus Boesch, v.boesch@qidenus.com

<sup>7</sup> Phorusgasse 8, 1040 Wien, Austria, <www.treventus.com>. Contatto: Ste-

phan Tratter, tratter@treventus.com

<sup>8</sup> Un accenno ai prodotti di 4Digital Books e Kirtas è rintracciabile in MICHAEL LESK, *Understanding Digital Libraries*, Second Edition, Amsterdam, Morgan Kaufmann, 2005, p. 64-65.

<sup>9</sup> Il software OCR di default è Abby Recognition Server.

#### Abstract

*Scanning hardware and software have come of age. Increasingly sophisticated equipment are coming to the market offering improved productivity and functionalities. Among the newest contraptions now available are automatic bookscanners. The Bayerische Staatsbibliothek in Munich (BSB) has recently staged a show of four of the most advanced examples of such equipment by four manufacturers: 4DigitalBooks, Kirtas Technologies, Qidenus, and Treventus. The article describes how the bookscanners showcased in Munich perform their job and offers a roundup of their basic features. The MDZ (Münchner Digitalisierungszentrum), which housed the event, is a state-of-the-art digitization centre attached to the BSB. It is equipped with a full range of scanners and has developed its own software with a view to streamline its operations. Today, thanks to continuous innovation and a dedicated team, the BSB can boast a total of 28.000 scanned books online.*