

Handbook on Cost Reduction in Digitisation

Roma, Ministero per i beni e le attività culturali, *Minerva Plus Project*, 2006, p. 36

Il processo di digitalizzazione è il processo attraverso il quale si traducono documenti analogici preesistenti nel modo reale (una fotografia, un disegno, un libro, un suono) in file digitali idonei ad essere visualizzati, ascoltati e processati da un elaboratore elettronico attraverso l'ausilio del software necessario. L'obiettivo è quello di riorganizzare la conoscenza in modo sempre più efficiente, rendendo fruibili documenti non sempre disponibili a tutti e razionalizzando il mare di notizie in un mondo ormai sommerso dalle informazioni.

Il processo di digitalizzazione, iniziato ormai da anni, ha metodi e costi non sempre accessibili.

Il volume realizzato da Simon Tanner nell'ambito del

Progetto Minerva dà opportuno rilievo alle problematiche legate ai costi della digitalizzazione e cerca di dare utili suggerimenti per la "riduzione dei costi nella digitalizzazione dei documenti". Il Progetto Minerva, iniziato nel marzo del 2002, ha lavorato per il coordinamento di politiche e programmi di digitalizzazione del patrimonio scientifico e culturale europeo, realizzando una rete di ministeri europei per la cultura. Fra le principali realizzazioni del Progetto Minerva vi sono appunto le Linee guida e le Raccomandazioni per la digitalizzazione del patrimonio culturale. Il progetto Minerva è uno degli esempi migliori di cooperazione tra strutture di diversi paesi dell'Unione europea.

Tanner affronta quello che oggi è uno dei problemi di maggior peso: il problema dei costi. Molti sono i progetti avviati con le migliori intenzioni che non vengono completati per mancanza di fondi.

Tanner presenta al lettore un quadro di insieme dei principali settori in cui può essere realizzato un taglio delle spese. Non a caso accenna, seppur fugacemente, al Book Search di Google come progetto di digitalizzazione a livello mondiale.

Il volume condensa in poche pagine le problematiche relative alla riduzione del costo del lavoro, alla automazione, alla selezione e preparazione dei documenti, al miglioramento dei risultati e al controllo di qualità.

Il costo del lavoro è quello che incide maggiormente nella realizzazione di un progetto. I problemi di maggior peso sono quelli legati alla qualità dell'intervento umano e alla qualità delle attrezzature che devono essere spes-

so sostituite, in quanto superate in qualità e risultati da sistemi tecnologici più avanzati. Tanner propone il frazionamento del lavoro in settori di competenza, utile per evitare il ricorso continuo a forze lavoro altamente specializzate ed estremamente costose. Per questi motivi egli esamina anche l'eventuale ricorso a "risorse esterne" (*outsourcing*), quando queste si dimostrino più affidabili soprattutto per la qualità del lavoro.

Un caso di ricorso all'*outsourcing* è quello della British Library per la realizzazione di due grandi progetti di digitalizzazione. I due progetti riguardavano l'archivio musicale (la digitalizzazione prevista era di 12.000 codificazioni audio per 3.900 ore di incisioni di suoni)¹ e i giornali pubblicati tra il 1800 e il 1900.² Entrambi i progetti, l'uno perchè richiedeva capacità tecniche particolari, l'altro per le dimensioni particolari dei giornali, richiedevano l'intervento di personale altamente specializzato e l'utilizzo di attrezzature dai costi molto elevati. La decisione della British Library di affidare il lavoro all'esterno, motivata dalle difficoltà di realizzazione, ha permesso il completamento del lavoro con costi inferiori a quelli che si sarebbero avuti se si fosse realizzato il progetto con risorse interne. L'ingresso dell'automazione nelle procedure di digitalizzazione è un processo abbastanza recente. L'automazione e i nuovi software hanno modificato e spesso ridotto l'intervento umano. È importante però riconoscere che l'automazione non è assolutamente la panacea per ridurre i costi, ma più che altro per accelerare la realizzazione dei progetti. Soprattutto il ricorso ad attrezz-

Figura 1

$$\text{Unit cost} = \frac{\text{capital expenditure}}{\text{number of items}} + \frac{\text{time expended per item}}{\text{number of items}}$$

zature più moderne ha consentito digitalizzazioni più veloci.

La digitalizzazione automatizzata se, da un lato, ha accelerato il meccanismo di digitalizzazione portandolo in alcuni casi addirittura a 3.000 pagine per ora, ha d'altra parte evidenziato le difficoltà di riconoscimento dei caratteri digitalizzati. La digitalizzazione automatizzata ha ampiamente ridotto i costi e i tempi nelle aree dei video e dei film, nell'ambito della digitalizzazione dei libri ha agevolato la cattura dei metadati, ma non sempre ha migliorato il riconoscimento del testo (OCR).³ Secondo una scala di valori individuata dall'autore, ma preesistente nel settore, i documenti sono classificabili nelle seguenti categorie: simple, noisy, complex, modern e historic.

Il caso esaminato da Tanner è il progetto realizzato dal Metadata Engine Project (METAe) per la Leopold Franzens Universität Innsbruck in Austria.⁴ Il progetto ha utilizzato un software di automazione che ha consentito la cattura dei metadati contemporaneamente alle procedure di digitalizzazione. Il progetto prevedeva la digitalizzazione di ALO (Austrian Literature Online) con libero accesso a più di 10.900 digitalizzazioni tra libri e giornali.

La selezione e la preparazione dei documenti da digitalizzare rappresentano settori del lavoro spesso ignorati, di importanza strategica per la riduzione dei costi.

Una corretta selezione del materiale potrebbe permettere una riduzione del costo totale del 20-30%. Alcuni suggerimenti sono quelli di se-

lezionare raccolte intere e non singole parti, selezionare originali della stessa natura fisica in modo da non dover cambiare in continuazione le attrezzature, ridurre al massimo i trasporti realizzando, nei limiti del possibile, digitalizzazioni *in loco*.

Tanner dedica poi spazio ai sistemi possibili per un aumento delle prestazioni. Particolare attenzione viene dedicata alle attrezzature. Le attrezzature di digitalizzazione sono costose, si svalutano prestissimo ed hanno un valore di rivendita molto basso anche dopo soli tre anni. Quindi, il suggerimento, molto semplice per altro, è quello di utilizzare al massimo e in tempi brevi le attrezzature acquistate per la digitalizzazione. L'autore elabora per questo la formula riportata in figura 1.

L'altro aspetto legato al miglioramento delle prestazioni è lo studio dei flussi di lavoro. L'identificazione delle priorità, l'identificazione delle attività da svolgere in parallelo e la scelta dei percorsi di digitalizzazione sono elementi che possono migliorare i flussi di lavoro e accelerare i processi di completamento dei progetti.

Il caso esaminato relativo all'aumento delle prestazioni è quello dell'Archivio nazionale svedese.⁵ Nell'ambito degli Archivi nazionali svedesi era attivo un centro di digitalizzazione denominato fino al 2003 SVAR. Tra il 2003 e il 2004 la sezione prese il nome di Media Konverterings Centrum (MKC);⁶ furono assegnate 77 unità di personale ed un budget di due milioni di euro. Il materiale da digitalizzare era di pro-

venienza e tipologia diversa, per cui il lavoro iniziale fu quello di dividere in settori tutto il materiale: documenti cartacei e microfilm di diversa provenienza. L'attività dell'MKC venne concentrata sulla produzione su larga scala di immagini digitali: il personale venne diviso su tre turni flessibili fino a coprire 20 ore giornaliere con una digitalizzazione di 90.000 immagini. Il materiale veniva affidato a gruppi diversi a seconda del formato e della diversa tipologia di lavoro: in tal modo il progetto è stato portato a termine nei tempi e con i costi prefissati.

Dovranno poi essere pianificate e messe in atto adeguate procedure di controllo di qualità da seguirsi durante la digitalizzazione, dalla fase di digitalizzazione delle immagini al riversamento delle immagini stesse e dei dati su supporto digitale. In particolare la qualità di tutti i materiali digitali dovrà essere garantita da diversi livelli di controllo, verifica e correzione: – verifica di leggibilità; – orientamento; – esposizione e contrasto; – corrispondenza dei colori; – altri parametri specifici delle immagini.

Il ruolo ricoperto da Simon Tanner – Direttore del KDCCS,

Digital Consultancy Services al King's College di Londra – gli ha consentito di seguire ben 450 progetti di digitalizzazione realizzati in Europa e in America. La "regola d'oro", come la definisce lui stesso, per la riduzione dei costi è focalizzare l'attenzione sui vari settori del lavoro, riducendo l'intervento umano sia durante i processi di digitalizzazione che durante la creazione dei metadati, selezionando con attenzione gli originali da digitalizzare, migliorando le prestazioni ed effettuando un continuo controllo di qualità. Il tutto ottenibile attraverso una attenta progettazione ed un rispetto preciso dei programmi da realizzare.

Maria Rosaria Bacchini

Coordinamento
biblioteche d'ateneo
Università degli studi di Napoli
mabacchi@unina.it

¹ Archival Sound Recording at the British Library, <<http://www.bl.uk/collections/sound-archive/archsoundrec.html>>.

² 19th Century British Newspapers at the British Library, <<http://www.bl.uk/collections/britishnewspapers1800to1900.html>>.

³ OCR = Optical Character Recognition.

⁴ <<http://meta-e.aib.uni-linz.ac.at>>.

⁵ <<http://www.ra.se/>>.

⁶ <<http://www.mkc.ra.se/>>.

