

VERONICA PESCE

*Per un archivio digitale.
Note in margine a una banca dati in costruzione*

In

I cantieri dell'italianistica. Ricerca, didattica e organizzazione agli inizi del XXI secolo.
Atti del XVII congresso dell'ADI – Associazione degli Italianisti (Roma Sapienza,
18-21 settembre 2013), a cura di B. Alfonzetti, G. Baldassarri e F. Tomasi,
Roma, Adi editore, 2014
Isbn: 9788890790546

Come citare:

Url = http://www.italianisti.it/Atti-di-Congresso?pg=cms&ext=p&cms_codsec=14&cms_codcms=581
[data consultazione: gg/mm/aaaa]

VERONICA PESCE

*Per un archivio digitale.**Note in margine a una banca dati in costruzione*

La comunicazione presenta il progetto di ricerca (finanziato nell'ambito del programma «Futuro in Ricerca 2012»): «Diffondere la cultura visiva: l'arte contemporanea tra riviste, archivi e illustrazioni. La storia dell'arte dalla fine dell'Ottocento agli anni Ottanta del Novecento vista attraverso fonti inesplorate, coniugando metodologie e sistemi di analisi multidisciplinari: critica storico-artistica, letteratura, semiotica, arti visive». L'attenzione si focalizza sul lavoro svolto dall'Unità di Genova, dove si sta procedendo alla digitalizzazione, insieme con i documenti dell'archivio «Mario Novaro», della rivista artistico-letteraria «La Riviera Ligure». Dare conto del lavoro avviato è qui occasione per riflettere criticamente sulle problematiche che si sono poste in fase di progettazione della banca dati e sulle possibilità in termini di fruizione, documentazione e raccolta di dati utili al lavoro critico che la piattaforma informatica offre, potenziando gli strumenti di lavoro tradizionali.

Queste note intendono proporre alcune riflessioni su ben precise e circoscritte applicazioni informatiche, relative allo studio delle scienze umanistiche, di cui sto avendo esperienza diretta nel Progetto di Ricerca che mi vede coinvolta. Il progetto è stato attivato e finanziato nell'ambito del programma ministeriale «Futuro in Ricerca 2012» e ha preso avvio nel marzo 2013. Il titolo del progetto è «Diffondere la cultura visiva: l'arte contemporanea tra riviste, archivi e illustrazioni». Il sottotitolo: «La storia dell'arte dalla fine dell'Ottocento agli anni Ottanta del Novecento vista attraverso fonti inesplorate, coniugando metodologie e sistemi di analisi multidisciplinari: critica storico-artistica, letteratura, semiotica, arti visive» precisa l'approccio transdisciplinare – dato molto valorizzato in questi programmi di ricerca, dunque elemento fondamentale su cui torneremo – che partendo dalla critica storico-artistica e letteraria si allarga ad altre discipline come la semiotica. Sono coinvolti quattro Atenei italiani (Scuola Normale Superiore di Pisa, con ruolo di coordinamento, Università di Siena, Università di Udine e Università di Genova). L'obiettivo è la realizzazione di una banca dati comune,¹ facente capo al Laboratorio Arti Visive della Scuola Normale Superiore di Pisa. Ogni unità di ricerca contribuisce con la digitalizzazione (e relative schedature) di archivi letterari e figurativi e/o di riviste letterarie e artistiche comprese in un ampio arco diacronico, da fine Ottocento agli anni '70 e '80 del Novecento. La banca dati è già disponibile in rete e consultabile; va da sé che essendo il lavoro ancora in corso non tutto il materiale è stato digitalizzato e l'interfaccia per l'utente insieme con le maschere di ricerca vengono migliorate e affinate man mano. Questa è forse la prima grande differenza con i supporti tradizionali: il mezzo informatico, insieme con l'ausilio della rete, consente di rendere immediatamente disponibile il lavoro, davvero in tempo in tempo reale.²

La nostra unità di ricerca, con il diretto coinvolgimento nel progetto di un Archivio privato genovese (la Fondazione Mario Novaro),³ si sta occupando della «Riviera Ligure» e del suo

¹ L'indirizzo web è: <http://www.capti.it/index.php?ParamCatID=1&id=2&lang=IT>. Si vedano in particolare le pagine dedicate alla presentazione generale del progetto e alla descrizione del lavoro dell'Unità di Genova.

² Questo è naturalmente un dato importantissimo, tanto più importante quanto più è ampia la mole del progetto e dunque quanto più il lavoro è lungo nel tempo. Si pensi a quale differenza farebbe per l'OVI (Istituto per l'Opera del Vocabolario Italiano) nella realizzazione del TLIO (*Tesoro della Lingua Italiana delle Origini*, titanico progetto per la realizzazione del vocabolario storico dell'italiano antico) se non fosse data la possibilità di avanzare parallelamente con lo spoglio delle fonti e la compilazione del vocabolario (che quindi non procede in ordine alfabetico) con la possibilità di ulteriori implementazioni via via. Cfr. <http://www.oivi.cnr.it/index.php?page=il-vocabolario>.

³ L'Ente si è costituito da trent'anni a questa parte (1983) in Fondazione o.n.l.u.s. per volontà degli eredi di Mario Novaro e di un gruppo di docenti dell'Ateneo genovese con lo scopo di valorizzare il lascito

archivio che raccoglie oltre 4000 documenti fra corrispondenza, autografi dei testi o realizzazioni artistiche poi pubblicati sul periodico che vide fra i suoi collaboratori personalità quali Giovanni Pascoli, Plinio Nomellini, Luigi Pirandello, Umberto Saba, Giuseppe Ungaretti. L'elenco potrebbe essere molto più lungo; ci siamo limitati a pochissimi nomi che già di per sé rendono conto delle ragioni che hanno portato alla scelta di questa rivista artistica e letteraria di altissimo profilo. È certamente nota la sua peculiare origine come bollettino pubblicitario della Ditta dell'Olio Sasso di Oneglia nel 1895 – è di fatto uno dei primi esempi di *House Organ* in Europa – e la successiva trasformazione a opera dal suo direttore e proprietario Mario Novaro, intellettuale, poeta, filosofo e imprenditore industriale, in rivista letteraria di prim'ordine, osservatorio privilegiato e incredibilmente avanzato sia della poesia italiana del primo Novecento sia dell'arte e della grafica aggiornata sullo stile Liberty, senza mai perdere la propria vocazione 'aziendale' e l'intento promozionale e pubblicitario.

Solo in un'esperienza culturale così poliedrica l'illustrazione di Plinio Nomellini all'*Inno all'Oliivo* di Giovanni Pascoli⁴ può trasformarsi nella celebre «musa all'olivo» del manifesto pubblicitario della ditta dell'Olio Sasso per divenire il logo poi riprodotto sulle etichette e sulle lattine dell'olio, in un mai più emulato connubio di arte, letteratura, comunicazione pubblicitaria e spirito aziendalistico. Una realtà di questo tipo non solo si presta, ma piuttosto esige un punto di osservazione effettivamente interdisciplinare che consenta di riunire e di incrociare costantemente dati diversi fra loro, dalla pubblicità, alla letteratura, alla grafica, più di quanto qualsiasi tipo di ricerca tradizionale consentirebbe.⁵ Senza dimenticare l'influenza di più ampio raggio che la rivista ha avuto contribuendo in modo determinate alla codificazione di una certa immagine della riviera ligure medesima: lo stesso titolo scelto per la testata va significativamente in questa direzione.⁶ Arte e letteratura sono state, almeno in parte, strumento per un intento pubblicitario e promozionale, esaltando le caratteristiche produttive, dunque socio-economiche e culturali della regione, favorendo la cristallizzazione di un paesaggio riconoscibile e peculiare, funzionale a fini propagandistici e oggi leggibile su un piano teorico attraverso la vastissima letteratura esistente, pure nel quadro delle finalità della *Convenzione*

intellettuale dell'imprenditore e direttore della «Riviera Ligure» e di promuovere iniziative a favore della cultura del Novecento in Liguria. Si veda il sito: <http://www.fondazionenovaro.it/new/index.php>. La collaborazione – certificata e siglata da precisi accordi scritti – di enti depositari di fondi archivistici e librari (ogni unità di ricerca ha attivato analoghe collaborazioni) si è rivelata indispensabile per concorrere al bando ministeriale.

⁴ G. PASCOLI, *Inno all'Oliivo*, «La Riviera Ligure», VII (1901), 30, 305-306.

Cfr.: <http://www.capti.it/index.php?ParamCatID=10&IDFascicolo=29&artgal=2&key=196&lang=IT>.

⁵ «La Riviera Ligure» vanta un interesse critico estremamente ricco e vivace, cresciuto continuamente negli anni; non sarà superfluo rimarcare che si sono occupati della rivista i maggiori studiosi del Liberty italiano, valga su tutti il nome di Rossana Bossaglia. Tuttavia la prassi scientifica e metodologica dei molti studi critici esistenti è stata sempre e soltanto di tipo settoriale: sono stati infatti analizzati separatamente gli aspetti letterari, artistici e addirittura economici e industriali della rivista. Fra i molti titoli: *La Riviera Ligure*, a cura di E. Villa e P. Boero, Treviso, Canova, 1975; P. BOERO, «La Riviera Ligure» fra industria e letteratura, Firenze, Vallecchi, 1984; R. BOSSAGLIA, *La Riviera Ligure un modello di grafica Liberty*, con un saggio di E. SANGUINETI, Genova, Costa & Nolan, 1985; *Lettere a «La Riviera Ligure». I. 1900-1905*, a cura di P. Boero, Roma, Edizioni di Storia e Letteratura, 1981; *Lettere a «La Riviera Ligure». II. 1906-1909*, a cura di P. Boero, Roma, Edizioni di Storia e Letteratura, 2001; *Lettere a «La Riviera Ligure». III. 1910-1912*, a cura di A. Aveto, P. Boero, F. Merlanti, Roma, Edizioni di Storia e Letteratura, 2003.

⁶ Si parte inizialmente con una volontà dichiarata di esaltare la Riviera di Ponente (come recita il titolo scelto inizialmente per la testata: «La Riviera Ligure di Ponente») e solo in un secondo momento la rivista diviene programmaticamente rappresentativa dell'intera riviera ligure, mutando la propria testata in «La Riviera Ligure». La precisazione «di Ponente» è mantenuta per le prime quattro annate, fino al 1898: «La Riviera Ligure di Ponente», IV (1898-99), 3. In questo primo periodo la rivista ha infatti un taglio dichiaratamente promozionale della regione e dei suoi prodotti, per poi aprirsi progressivamente a collaborazioni di rilievo sia in campo letterario sia in campo artistico. Basti scorrere gli indici del periodico per constatare la graduale apertura: P. BOERO, «La Riviera Ligure» (1895-1919). *Indici*, Milano, All'insegna del pesce d'oro, 1976.

Europea del Paesaggio che definisce il paesaggio medesimo anche in funzione dell'identità della popolazione, quale patrimonio al contempo naturale e culturale da salvaguardare ad ogni livello.⁷

Se è vero quindi che un approccio non tradizionale e pluridisciplinare non può che giovare alla non infrequente complessità degli oggetti dei nostri studi e delle nostre ricerche, focalizziamo ora la nostra attenzione sugli aspetti tecnico-metodologici del progetto. Il primo punto – forse anche quello più problematico – è la digitalizzazione. Ci si potrà chiedere cosa significa digitalizzare un documento, quali scelte siano necessarie e in funzione di quale scopo, ma forse dovremmo prima di tutto rispondere alla domanda: «perché digitalizzare?». E magari interrogarci sulle motivazioni per cui si danno tutte queste proposte progettuali di digitalizzazione e quindi quali ragioni abbia un investimento così forte a livello politico-culturale, di cui il programma *Futuro in Ricerca* è caso emblematico in Italia, ma non fanno eccezione i progetti europei, in particolar modo gli obiettivi di *Horizon 2020*.⁸

Fra le motivazioni che muovono alla digitalizzazione si suole parlare di conservazione. Ma a ben vedere la finalità conservativa è l'aspetto più problematico. È noto infatti che la tecnologia informatica ha uno sviluppo velocissimo la cui ineluttabile conseguenza è un altrettanto veloce invecchiamento sia dei linguaggi utilizzati per immagazzinare le nostre informazioni e costruire le nostre banche dati sia dei supporti materiali (il cosiddetto *hardware*) quindi le memorie su cui quegli stessi dati vengono immagazzinati. Il problema non va certo minimizzato. Perché e per chi facciamo questi lavori se fra dieci o forse vent'anni – e forse sbaglio per eccesso – sarà superata la tecnologia che li ha prodotti e (ancor più grave!) sarà superata la tecnologia che può leggerli e fruirne? Qualche umanista della vecchia guardia potrà constatare non senza compiacimento che il codice Hamilton 1470 di Berlino o il Vaticano latino 3195 hanno a oggi una vita di gran lunga più duratura della loro eventuale digitalizzazione; ma senza andare a scomodare autografi e idiografi illustri, basti prendere atto che la vita della carta è di norma più lunga di quella di una qualsiasi memoria informatica e va meno soggetta a perdite accidentali.⁹

La preservazione dei documenti costituisce quindi un aspetto estremamente problematico e apparentemente senza via d'uscita. Esso è tuttavia compensato da altri importanti fattori, cioè dall'allargamento della fruibilità dei materiali e dalla possibilità di avere a disposizione nuovi strumenti di consultazione dei documenti stessi e quindi di ricerca. Poter fruire di documenti conservati a chilometri di distanza e magari difficilmente accessibili per molteplici ragioni di carattere conservativo, gestionale etc., va da sé, può semplificare e di molto il lavoro di chi fa ricerca, anche se la fruizione in rete non può mai essere sostitutiva dello studio del documento originale, torneremo nello specifico anche su questo punto.

Alla quantità di materiali di studio che continuamente si riversano in rete si lega tuttavia un altro problema complementare, ossia la gestione di tutto il materiale che di progetto in

⁷ Già il *Preambolo* dichiara che «il paesaggio rappresenta una componente fondamentale del patrimonio culturale e naturale dell'Europa, contribuendo così [...] al consolidamento dell'identità europea». Si legge poi all'articolo 5 comma a che «Ogni parte si impegna a riconoscere giuridicamente il paesaggio in quanto componente essenziale del contesto di vita delle popolazioni, espressione della diversità del loro comune patrimonio culturale e naturale e fondamento della loro identità». Le citazioni sono tratte dalla traduzione del testo ufficiale in inglese e in francese della *Convenzione Europea del Paesaggio* (Firenze, 20 ottobre 2000), curata da M. R. Guido e D. Sandroni dell'Ufficio Centrale per i Beni ambientali e paesaggistici. Cfr. <http://www.convenzioneeuropeapaesaggio.beniculturali.it/>.

⁸ Si verifichi direttamente sui rispettivi bandi di concorso, usciti negli anni scorsi o attualmente aperti: <http://futuroinricerca.miur.it/> e <http://ec.europa.eu/programmes/horizon2020/>.

⁹ Questi aspetti sono ampiamente trattati nella letteratura esistente su questi temi. Rimandiamo fra gli altri a: AA.VV., *Informatica per le scienze umanistiche*, a cura di T. Numerico e A. Vespignani, Bologna, il Mulino, 2003; F. TOMASI, *Metodologie informatiche e discipline umanistiche*, Prefazione di D. Buzzetti, Roma, Carocci, 2008; Più rivolti alla riflessione sui mutamenti legati alla filologia e alla scrittura: D. FIORMONTE, *Scrittura e filologia nell'era digitale*, Torino, Bollati Boringhieri, 2003; P. ITALIA, *Editing Novecento*, Roma, Salerno Editrice, 2013.

progetto viene via via schedato, catalogato, marcato, digitalizzato e spesso lanciato nel *mare magnum* del web dove paradossalmente – ma si tratta di un paradosso di internet intero e i nostri progetti di ricerca non fanno eccezione – resta a volte sconosciuto o conosciuto solo in minima parte. L'idea guida dell'ADI e del presente convegno vanno già nella giusta direzione con l'esplicita proposta di iniziare una riflessione sulla necessità di un coordinamento della ricerca in genere. Il discorso dovrebbe approfondirsi anche in relazione a questi progetti, sempre più parte integrante dell'attività accademica, ma bisognosi di occasioni di relazione e di confronto, non solo in vista di un auspicabile, ma sicuramente non semplice, coordinamento nelle iniziative pure in questo campo, ma anche e in primo luogo nel tentativo di mettere in comunicazione i progetti fra loro, censirli, recensirli, come si fa tradizionalmente per le pubblicazioni scientifiche.

Ma torniamo per un momento al punto da cui siamo partiti: la digitalizzazione. Cosa significa digitalizzare un documento?

Digitalizzare un'immagine (una fotografia, la pagina di un manoscritto, un dipinto ecc.) significa riprodurre l'oggetto materiale, cioè la rappresentazione analogica, in formato digitale, secondo parametri variabili a seconda dello scopo della digitalizzazione e quindi della destinazione d'uso del file realizzato. [...] L'obiettivo finale è però sempre la conversione di un segnale analogico in una forma digitale.¹⁰

Ma quali immagini offriremo all'utente? Con quali caratteristiche? Di alta o di bassa qualità? Ovviamente non esiste una risposta 'giusta', ma la ricerca costante di un punto di equilibrio fra diverse esigenze. L'uso di un formato di salvataggio compresso (JPEG) o privo di compressione (TIFF) cambia radicalmente il 'peso' dell'immagine, quindi lo spazio di memoria necessario, su cui incide pure la risoluzione dell'immagine, ossia i *dpi* (*dots per inch*, ovvero 'punti per pollice');¹¹ è diversa la velocità e dunque il tempo necessario per caricare il file in rete da parte di chi lavora al progetto ed è diverso poi il tempo necessario per scaricare o anche solo visualizzare il file da parte dell'utente. E la differenza non è solo in termini di tempo: entra in gioco una rilevante questione qualitativa. Con la compressione si perdono 'informazioni' la cui mancanza non consentirà alcune importanti manipolazioni dell'immagine, per esempio un ingrandimento di qualità accettabile o altri interventi finalizzati a migliorare la leggibilità dei documenti medesimi. Si tratta di operazioni elementari che possono tuttavia costituire un utilissimo e prezioso potenziamento rispetto agli strumenti tradizionali, soprattutto nei casi in cui si può agevolare la lettura proprio a un mero livello di decifrazione del testo:

Una foto digitale ad alta definizione permette di ingrandire decine di volte l'immagine rispetto alla sua dimensione naturale, di ruotarla, di modificarne colore e contrasto cromatico, e addirittura, se visualizzata con particolari programmi di trattamento delle immagini, di leggere sotto cassatura ciò che a occhio nudo risulta indecifrabile.¹²

D'altro canto un'immagine compressa proprio per la sua scarsa qualità difficilmente potrà essere utilizzata per una riproduzione a stampa. Il dato è rilevante se si ragiona di 'protezione' (*copyright*) delle immagini che si rendono disponibili in rete. Non è data tuttavia una soluzione che metta del tutto al riparo da possibili riproduzioni non volute. Una parziale via d'uscita consiste nell'impiego della cosiddetta filigrana digitale, il *watermark*: un 'marchio' che si sovrappone all'immagine e che sicuramente ne garantisce un maggiore livello di protezione, ma allo stesso tempo può disturbare la lettura.

Giunti quindi a definire un compromesso accettabile tra qualità, livello di 'protezione' e velocità di (s)caricamento dell'immagine, il file con la riproduzione del documento è pronto per essere schedato e caricato in rete. Come affermano gli specialisti: «è essenziale che una

¹⁰ F. TOMASI, *Metodologie informatiche...*, 169.

¹¹ Per approfondimenti tecnici su questo aspetto, cfr. ancora ivi, 172 e sgg.

¹² ITALIA, *Editing...*, 216.

collezione di immagini digitali non sia una raccolta non descritta e non strutturata». ¹³ Le immagini andranno correlate a una serie di informazioni, dette metadati, relative all'immagine stessa, alle caratteristiche fisiche dell'originale riprodotto e ai suoi contenuti. Questa operazione di 'schedatura' è una parte a dir poco fondamentale sulla cui base andrà orientata fin dal principio la scelta progettuale della banca dati. Si dovrà distinguere tra le informazioni che vogliamo siano semplicemente visualizzate dall'utente relative a ogni specifico documento (come una didascalia) e i dati che intendiamo siano ricercabili dall'utente e interrogando il database restituiscano determinati risultati. La banca dati dovrà prevedere appositi 'campi' da compilare al momento del caricamento in rete e per ciascuno di essi, con una diversa marcatura, si avrà un campo ricercabile o meno. Talvolta i dati andranno ripetuti a livelli differenti: per indicare il numero del fascicolo, se intendiamo schedare e caricare in rete un periodico come nel caso di «Riviera Ligure», si compilerà un campo numerico e uno alfanumerico. Il primo sarà funzionale all'architettura del database e consentirà il giusto ordinamento dei fascicoli, il secondo restituirà all'utente l'informazione così come si legge in originale sulla copertina (in numero, in lettere etc.). Discorso analogo per altre informazioni: il campo autore, per proporre ancora un esempio, dovrà essere sempre uniforme (prima cognome e poi nome) per consentire l'identificazione nel database e la corretta restituzione all'utente, si darà poi un campo firma dove il nome sarà trascritto con criterio filologico così come appare sull'originale, con eventuale pseudonimo o semplicemente non scritto per esteso. Allo stesso modo non sono schedati solo i testi, ma pure le illustrazioni, la pubblicità, le decorazioni grafiche, i paratesti. Tutti elementi che diventano tanto più significativi se messi in relazione gli uni agli altri: la nascita di una precisa idea grafica realizzata sulla rivista, seguita attraverso la corrispondenza, la menzione di un testo letterario nelle lettere o il suo autografo legati al fascicolo di «Riviera» in cui è pubblicato: ecco che la ricerca dell'*Inno all'Olivo* di Giovanni Pascoli darà come risultato il rimando alla pubblicazione su periodico, l'autografo del testo, le lettere ad esso legate e l'illustrazione di Plinio Nomellini. ¹⁴

Le potenzialità in questo senso sono notevoli. Naturalmente più informazioni inseriamo e più relazioni creiamo fra i dati stessi, più l'architettura della banca dati si fa complessa e più si fa gravoso il lavoro per chi costruisce il database e per chi immette i dati da un lato e per l'utente che dovrà fare ricerca dall'altro, perché saranno più lenti i tempi di restituzione dei risultati. L'attenzione nel processo di schedatura andrà anche tesa a evitare quello che in gergo si chiama sinestesicamente «rumore», ossia una ridondanza di risultati non richiesti all'atto della ricerca da parte dell'utente. Di qui l'esigenza di precisare il più attentamente possibile quali siano le informazioni essenziali da restituire agli utenti, delimitando altrettanto precisamente i campi ricercabili.

Ancora una precisazione sulla fruizione in rete dei documenti digitalizzati: poniamo una lettera di Giovanni Pascoli o l'autografo dell'*Inno all'Olivo* già ricordato. Parafrasando René Magritte che come noto scriveva sotto il dipinto di una pipa iperrealista: «Questa non è una pipa» (1948) potremmo mettere una didascalia al nostro file: «Questo non è l'autografo di Giovanni Pascoli». E non è solo una provocazione. Bisogna sempre aver chiaro che digitalizzare è tutto fuorché un'operazione neutra che fa passare un oggetto materiale, come un manoscritto o la pagina di una rivista, per una complessa operazione di codifica che la riproduce e ne restituisce un'immagine virtuale. L'utente non ha più davanti fisicamente il documento e di questo si dovrà tenere conto costantemente. Poniamo il caso di una lettera redatta su un foglio ripiegato a formare due carte a loro volta scritte ciascuna sul recto e sul verso magari in differenti sensi di scrittura. Come sarà visualizzato dall'utente? Si può restituire in qualche modo la 'materialità' dell'oggetto che stiamo riproducendo in rete? Come salvare leggibilità e rigore filologico? E non ultimo che cosa interessa ed è più utile al fruitore di queste banche dati?

L'utente nel caso preso a campione vedrà quattro immagini, risultato di quattro differenti scansioni. E in quale ordine? Quello di lettura o nell'ordine in cui si presentano nella loro collocazione archivistica? Fermo restando che sarà necessaria una precisa descrizione del

¹³ TOMASI, *Metodologie informatiche...*, 182.

¹⁴ Cfr.: <http://www.capti.it/search.php?s=inno+all%27olivo&kindofasearch=2&lang=IT>.

documento, il tentativo di restituire fedelmente l'originale (operazione a rigore 'corretta') qui potrebbe perdere senso perché sarebbe solo una simulazione, un tentativo di restituire qualcosa che nel processo di digitalizzazione è già diventato una sequenza binaria che poi la macchina ha ricomposto sul nostro schermo, conservando più e meno informazioni a seconda del formato e delle impostazioni che sono state definite in partenza. Su questa linea potremmo continuare enumerando una serie di altri problemi collaterali, per esempio l'opportunità di caricare il documento nella sua interezza, fogli non scritti compresi, che costano spazio in termini di memoria e allungano (inutilmente?) i tempi di caricamento.

Si abbia poi sempre presente che fatta una scansione abbiamo un'immagine digitale, non un testo digitale, riconosciuto come tale dal computer. Quindi perché l'immagine, laddove rappresenta un documento testuale, abbia un significato e non sia 'muta' nel suo contenuto, dando la possibilità di estrarne le relative informazioni, va in qualche modo elaborata affinché sia riconosciuta come testo dal computer. Per questo esiste il Sistema di riconoscimento ottico dei caratteri, detto OCR (*Optical Character Recognition*), capace di riconoscere il testo: cioè riconosce il 'tracciato' (o meglio un tracciato tipo) del carattere alfanumerico. L'OCR è la base per poter trasformare un'immagine in 'testo' rendendo così ricercabili le parole. Esso diventa dunque uno strumento preziosissimo alla base di qualsiasi analisi linguistica sui documenti digitalizzati. È il cosiddetto *text retrieval* in grado di restituire in pochi secondi le forme testuali (per brevità possiamo parlare di parole) presenti in un testo, dunque si potranno rintracciare occorrenze lessicali, elaborare liste di frequenza, vere e proprie concordanze ecc. Arrivare ad un buon OCR però è ancora operazione piuttosto lunga perché i software, anche i migliori, non danno risultati pienamente affidabili e hanno margini di errori piuttosto alti soprattutto sui caratteri accentati o semplicemente sulla pagina più deteriorata. L'impiego dell'OCR esige pertanto attente revisioni; il risultato finale è prezioso ma costa un dispendio altissimo di tempo e risorse.

Rispetto al progetto qui presentato un bilancio è senz'altro prematuro. Può però procedere di pari passo al nostro lavoro la riflessione critica su questi strumenti, tanto più significativa ed efficace in corso d'opera. Certo non mancano gli aspetti problematici ma danno altrettante sollecitazioni le potenzialità che questi strumenti offrono ma che dovranno passare poi al vaglio della pratica e dell'uso. Per ora siamo impegnati nella ricerca costante di un punto di equilibrio fra istanze diverse: protezione *versus* fruibilità, leggibilità *versus* filologia, costi *versus* risultati etc. Certamente una sfida che impone di ripensare i tradizionali statuti dei nostri studi; una sfida che oggi anche le discipline umanistiche devono saper raccogliere.