

## Risorse digitali e strutture di dati

Dino Buzzetti

Una rassegna delle risorse digitali disponibili in rete riguardanti la filosofia medievale mostrerebbe che la grande maggioranza dei materiali accessibili è costituita da insiemi di pagine del WorldWideWeb, il sistema informativo distribuito, a struttura ipertestuale, sviluppatosi in Internet. E tale situazione è certamente prevalente anche negli altri settori degli studi medievalistici. Ora, è legittimo chiedersi se, e in che misura, un «insieme di pagine Web», possa effettivamente «promuovere l'uso di tecniche informatiche nello studio della storia medievale». È in questo modo, infatti, che Andrea Zorzi, in un recente intervento, presenta il progetto di *Reti Medievali*, i cui promotori si pongono programmaticamente l'obiettivo di «modificare l'opinione prevalente circa il basso livello scientifico delle pubblicazioni elettroniche».<sup>1</sup> Certamente, per procedere in questa direzione, occorre interrogarsi sullo «statuto dell'editoria scientifica digitale», come si propone di fare Rolando Minuti affrontando, sulla «Rivista» di *Reti Medievali*, il «problema della pubblicazione elettronica».<sup>2</sup>

Ma il problema si riduce semplicemente ad un problema di accreditamento della forma non cartacea di pubblicazione? Si tratta semplicemente di rivendicare il «riconoscimento legale» della «pubblicazione elettronica» e di equiparare a pieno titolo alla «pubblicazione cartacea» tradiziona-

<sup>1</sup> A. ZORZI, *'Reti Medievali': An initiative of new communication of historical knowledge*, intervento al Convegno internazionale 'Scholarly Communication and Academic Presses', Firenze, 22 marzo 2001, in *Reti medievali*, sez. «Novità», <URL: <http://www.rm.unina.it/novita/Nov-presenta-FI.htm>>.

<sup>2</sup> R. MINUTI, *Le incognite della 'pubblicazione' on-line: contributo all'avvio di una discussione*, in *Reti medievali*, sez. «Rivista», 2, 2001, <URL: [http://www.storia.unifi.it/\\_RM/rivista/forum/Minuti.htm](http://www.storia.unifi.it/_RM/rivista/forum/Minuti.htm)>.

le la pubblicazione in forma non cartacea della *stessa* opera, o non occorre piuttosto chiedersi se la forma digitale non realizzi una qualità *diversa* della produzione intellettuale? La questione è tutt'altro che oziosa; infatti, evitando di affrontare questo problema, difficilmente si riuscirà a spiegare perché «il meccanismo della comunicazione della ricerca», in campo umanistico, «continui a seguire le strategie tradizionali». La questione della forma di pubblicazione, cartacea o digitale, è secondaria rispetto al problema della qualità scientifica della produzione intellettuale. Il riconoscimento legale della pubblicazione digitale non può essere considerato condizione sufficiente ad accreditare pubblicazioni digitali di scarso valore scientifico. Il riconoscimento legale è certamente una condizione necessaria per l'accREDITAMENTO della produzione intellettuale in forma digitale, ma non può essere considerato in alcun modo condizione sufficiente. La sola possibilità di accesso attraverso i «nuovi scenari» propri «del 'mondo nuovo' della comunicazione» non garantisce da sé la qualità di una pubblicazione digitale. Né può garantirla il suo riconoscimento legale. Dunque non pare che la discussione sulla *qualità* delle risorse digitali e sulla *specificità* della forma digitale di produzione intellettuale possa essere assolutamente evitata. A meno che non si miri semplicemente ad ottenere «efficacia e rapidità nella diffusione» e non si voglia circoscrivere il problema alla sola «comunicazione» dei risultati della ricerca. Nella sostanza, sotto questa luce, dovremmo affrontare una questione eminentemente biblioteconomica, considerando esclusivamente i problemi della «conservazione» e della «diffusione» dei «documenti elettronici»,<sup>3</sup> senza affrontare la questione propriamente scientifica dell'adeguatezza metodologica della produzione intellettuale in forma digitale.

Il vero interrogativo che lo sviluppo delle nuove tecnologie dell'informazione pone alla comunità scientifica pare allora essere il seguente: può la forma digitale di rappresentazione della conoscenza mutare la qualità stessa dei risultati della ricerca? Fin dal primo delinearci del nuovo campo di studi, oggi chiaramente riconosciuto e comunemente descritto come 'informatica umanistica,' Tito Orlandi sosteneva:

vorremmo sottolineare che non è tanto la quantità di dati resa disponibile dalle nuove tecnologie che conta, quanto la loro formalizzazione, imposta dall'informatica, e l'uso di procedimenti automatici, consentiti dall'informatica.

<sup>3</sup> *Ibid.*

Sicché, «si deve riflettere soprattutto sull'influsso» che i procedimenti informatici possono avere «sullo sviluppo e sulla metodologia» delle discipline umanistiche.<sup>4</sup> E c'è più di un fondato motivo per sostenere che la forma digitale di rappresentazione della conoscenza può mutare i metodi e i procedimenti della ricerca umanistica e può permettere di risolvere problemi difficilmente superabili «lungo i binari stabiliti dalle forme tradizionali» dell'indagine scientifica.<sup>5</sup> Per non citare che un solo esempio concreto, restando nel campo degli studi medievali, l'edizione di testi manoscritti a tradizione spiccatamente 'fluida,' collazionabile direttamente soltanto in parte, pare potersi realizzare adeguatamente solo nella forma di un *database* contenente la trascrizione completa di tutti i testimoni. L'informatica offre, in questo caso, soluzioni non praticabili adottando altre forme di rappresentazione del testo e permette di affrontare difficoltà che in altro modo non potrebbero essere concettualmente o concretamente superate.<sup>6</sup> Inoltre, le recenti discussioni riguardanti i modelli strutturali delle «biblioteche digitali»<sup>7</sup> rendono del tutto evidente che le «modalità di conservazione e di accesso»<sup>8</sup> ai materiali archiviati dipendono

<sup>4</sup> T. ORLANDI, *Informatica umanistica*, Roma, La Nuova Italia Scientifica, 1990, p. 12.

<sup>5</sup> MINUTI, *Le incognite della 'pubblicazione' on-line*, cit.

<sup>6</sup> Cfr. D. BUZZETTI - A. TABARRONI, *Informatica e critica del testo: il caso di una tradizione 'fluida'*, in «Schede umanistiche», n.s., 1, 1991, n. 2, pp. 185-193; D. BUZZETTI, *Image Processing and the Study of Manuscript Textual Traditions*, in «Historical Methods», 28, 1995, 3 f., pp. 145-154; D. BUZZETTI - A. TABARRONI, *Database Edition of Non-collatable Textual Traditions*, in *The Electric Scriptorium: Electronic Approaches to the Imaging, Transcription, Editing, and Analysis of Medieval Manuscript Texts*, A Physical and Virtual Conference (Calgary, 10-12 November 1995), <URL: <http://www.ucalgary.ca/~scriptor/papers/buzzetti.html>>; D. BUZZETTI, *Il testo 'fluido': Sull'uso dell'informatica nella critica e nell'analisi testuale*, in *Filosofia & informatica*, Atti del primo incontro italiano sulle applicazioni informatiche e multimediali nelle discipline filosofiche (Convegno Nazionale della Società Filosofica Italiana: Roma, 23-24 novembre 1995), a cura di Luciano Floridi, Torino, Paravia, 1996, pp. 85-93; D. BUZZETTI - M. REHBEIN, *Textual Fluidity and Digital Editions*, in *Text Variety in the Witnesses of Medieval Texts*, Proceedings of the International Workshop, Sofia, 21-23 September 1997, edited by Milena Dobreva, Sofia, Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, 1998, pp. 14-39.

<sup>7</sup> Cfr. W. Y. ARMS, *Key Concepts in the Architecture of the Digital Library*, in «D-Lib Magazine», 1:1, 1995, <URL: <http://www.dlib.org/dlib/July95/07arms.html>>; C. LAGOZE, *The Warwick Framework: A Container Architecture for Diverse Sets of Metadata*, in «D-Lib Magazine», July/August 1996, 2:7, 1996, <URL: <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>>.

<sup>8</sup> MINUTI, *Le incognite della 'pubblicazione' on-line*, cit.

dalla natura stessa degli «oggetti digitali»<sup>9</sup> creati direttamente dagli studiosi e non possono essere assolutamente considerate come puri «aspetti tecnici»,<sup>10</sup> che converrebbe «delegare ad uno *staff* editoriale specializzato».<sup>11</sup>

Il giudizio dello studioso è essenziale nella produzione stessa dell'oggetto digitale. Solo lo studioso, e non il tecnico, è in grado di valutare la qualità scientifica di una risorsa digitale. Infatti sono «solo le esigenze analitiche della ricerca», quelle che «possono fornire, oltre alle motivazioni euristiche, i *criteri di adeguatezza* per un'eventuale rappresentazione digitale» dei materiali di studio. Se si può così fondatamente sostenere che un'edizione digitale «si giustifica pienamente solo se può fornire soluzioni a problemi difficilmente affrontabili in modo diverso»,<sup>12</sup> allora si può altrettanto legittimamente affermare che «la struttura della rappresentazione digitale diviene rilevante» per valutare l'«adeguatezza» di qualsivoglia materiale di studio archiviato in *machine readable form*.<sup>13</sup> È dunque la *struttura* della forma di rappresentazione digitale ciò che costituisce l'elemento principale per valutarne l'adeguatezza. E prima ancora che per l'archiviazione di dati a struttura più complessa, come quelli richiesti per la descrizione di fonti materiali, o per la registrazione e l'elaborazione statistica di conoscenze di fatto, queste considerazioni risultano del tutto evidenti per la stessa rappresentazione digitale del testo.

Da un punto di vista informatico, l'informazione testuale viene rappresentata mediante la codifica binaria dei caratteri e il testo viene definito semplicemente come «informazione codificata come caratteri o come sequenze di caratteri». Sicché, da un punto di vista informatico, per testo non si intende il «materiale letterario nella forma in cui viene originaria-

<sup>9</sup> R. KAHN - R. WILENSKY, *A Framework for Distributed Digital Object Services*, <URL: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>>.

<sup>10</sup> MINUTI, *Le incognite della 'pubblicazione' on-line*, cit.

<sup>11</sup> ZORZI, *Reti Medievali: An initiative*, cit.

<sup>12</sup> D. BUZZETTI, *Ambiguità diacritica e Markup: Note sull'edizione critica digitale*, in *Soluzioni informatiche e telematiche per la filologia*, a cura di S. Albonico, Atti del Seminario di studi (Pavia, 30-31 marzo 2000), Pavia, Università degli Studi di Pavia, 2000 (Pubblicazioni telematiche, n. 1), <[http://dobc.unipv.it/diplamm/Pubblicazioni-telematiche/dino\\_buzzetti.htm](http://dobc.unipv.it/diplamm/Pubblicazioni-telematiche/dino_buzzetti.htm)>, § 1.

<sup>13</sup> ID., *Rappresentazione digitale e modello del testo*, in *Il ruolo del modello nella scienza e nel sapere*, Atti del Convegno (Roma, 27-28 ottobre 1998), Roma, Accademia Nazionale dei Lincei, 1999 (Contributi del Centro Linceo Interdisciplinare Beniamino Segre, 100), p. 129.

mente scritto da un autore». <sup>14</sup> È però del tutto evidente che non è possibile esprimere tutta l'informazione contenuta in un documento impiegando una pura e semplice successione di caratteri e che occorre aggiungere ulteriori informazioni per distinguere al suo interno le sequenze che costituiscono, per esempio, i titoli, o le note, o i riferimenti marginali, come i numeri di pagina. In breve, la nozione informatica del testo non permette di rappresentare in modo esaustivo tutta l'informazione trasmessa da un documento. L'«esaustività della rappresentazione» costituisce dunque una prima fondamentale condizione di adeguatezza. <sup>15</sup> Il *markup*, ossia l'aggiunta di marcatori (*tags*) all'interno della sequenza di caratteri che costituisce la rappresentazione informatica del testo, è stato giudicato lo strumento idoneo per risolvere questo problema e lo studioso che voglia valutare l'adeguatezza della rappresentazione digitale di materiali testuali deve di conseguenza prestare attenzione alle conseguenze prodotte dall'inserimento dei marcatori entro la successione dei caratteri. Poiché l'inserimento dei marcatori costituisce «una delle diverse tecniche possibili per rappresentare la struttura» di un «documento», <sup>16</sup> applicare al testo una forma di *markup* significa assegnare al testo una struttura. Con l'aggiunta dei marcatori, la struttura della rappresentazione informatica del testo non risulta più costituita da una semplice successione di caratteri, ma da blocchi di caratteri posti in relazione, in un certo modo, gli uni con gli altri.

La *struttura di dati* che ne deriva deve però soddisfare un'altra fondamentale condizione di adeguatezza e si tratta di nuovo di una condizione di adeguatezza che solo lo studioso può essere in grado di valutare. La struttura di dati che costituisce la rappresentazione informatica del testo «deve essere funzionale alle operazioni analitiche» che il suo studio richiede, ossia, in generale, alle «operazioni critiche di costituzione o di interpretazione del testo». <sup>17</sup> Si consideri, per esempio, la pratica adottata e ampiamente seguita dalla comunità degli studiosi per garantire una forma di rappresentazione digitale del testo filologicamente corretta, una pratica

<sup>14</sup> C. DAY, *Text Processing*, Cambridge, Cambridge University Press, 1984, p. 1.

<sup>15</sup> BUZZETTI, *Rappresentazione digitale e modello del testo*, p. 128.

<sup>16</sup> D. R. RAYMOND - F. W. TOMPA - D. WOOD, *Markup Reconsidered*, paper presented at the First International Workshop on Principles of Document Processing, Washington DC, October 22-23, 1992 <URL: <http://www.csd.uwo.ca/staff/drraymon/.papers/markup.ps>>, Abstract, p. 1.

<sup>17</sup> BUZZETTI, *Rappresentazione digitale e modello del testo*, p. 129.

fondata sulle *Guidelines* della Text Encoding Initiative (TEI),<sup>18</sup> ossia di quella che può a pieno titolo essere considerata «la più organica proposta finora avanzata» per la «codifica dei testi» e per l'«interscambio dei dati nel campo della ricerca umanistica».<sup>19</sup> L'adozione delle norme di codifica raccomandate dalla TEI, «che si propongono, ormai, come punto di riferimento ineludibile per chi si occupa di trattamento informatico dei testi»,<sup>20</sup> ha avuto, tra le sue conseguenze, quella di sollevare il perentorio interrogativo *Che cos'è il testo, in realtà?*<sup>21</sup> e di suggerire un'altrettanto perentoria risposta. Infatti, considerando come 'oggetto dotato di contenuto' una porzione di documento che contiene o è contenuta in altri oggetti dotati di contenuto, o porzioni del documento, e che forma con essi una 'gerarchia' di relazioni di contenimento, i cui elementi minimi sono 'ordinati' in successione nella sequenza di caratteri di cui il documento si compone, il testo è stato definito

come 'gerarchia ordinata di oggetti dotati di contenuto' ovvero come 'OHCO' (*Ordered Hierarchy of Content Objects*).<sup>22</sup>

Ma può tale definizione, o la 'tesi OHCO' come è stata chiamata, essere veramente assunta, per restare alle intenzioni dei suoi autori, come «modello fondamentale del testo»?<sup>23</sup> La definizione è certamente corretta da un punto di vista informatico, ma ci si deve chiedere se tale definizione formale della struttura di dati che fornisce la rappresentazione digitale del testo, una definizione fondata sulle raccomandazioni della TEI, e sulla scelta dello Standard Generalized Markup Language (SGML) come esclusivo linguaggio di codifica, soddisfi effettivamente le esigenze del lavoro critico degli studiosi. E non è certo sorprendente constatare che l'attività stessa di marcatura condotta in conformità con le norme della TEI

<sup>18</sup> C. M. SPERBERG-MCQUEEN - L. BURNARD (eds.), *Guidelines for Text Encoding and Interchange*, Chicago and Oxford, Text Encoding Initiative, 1994.

<sup>19</sup> L. BURNARD, *An Introduction to the Text Encoding Initiative*, in D. Greenstein (ed.), *Modelling Historical Data*, St. Katharinen, Max-Planck-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, 1991, p. 83.

<sup>20</sup> G. GIGLIOZZI, *Il testo e il computer: Manuale di informatica per gli studi letterari*, Milano, Bruno Mondadori, 1997, p. 109.

<sup>21</sup> Così suona il titolo di un noto e ampiamente discusso intervento: cfr. S. J. DEROSE - D. D. DURAND - E. MYLONAS - A. H. RENEAR, *What is Text, Really?*, in «Journal of Computing in Higher Education», 1:2, 1990, pp. 3-26.

<sup>22</sup> *Ibid.*, p. 6.

<sup>23</sup> *Ibid.*

abbia fatto emergere «alcuni problemi pratici che sembrano mettere questa tesi in discussione».<sup>24</sup> È stato evidentemente il giudizio degli studiosi a far nascere l'esigenza di apportare modifiche alla nuova nozione esclusivamente informatica del testo. Non poteva certo risultare soddisfacente una concezione del testo coincidente con la definizione della struttura di dati giudicata più idonea ad offrirne la rappresentazione digitale. Il testo è molto di più di una mera sequenza di caratteri strutturata mediante l'inserimento di marcatori e non è possibile far dipendere la concezione del testo dalla sintassi di un determinato linguaggio di codifica. Si era passati dall'idea della pura e semplice stringa di caratteri priva di marcatori all'idea della stringa di caratteri strutturata per mezzo di marcatori, ma si restava ancorati ad un tipo di definizione che assimilava il testo alla struttura di dati usata per rappresentarlo, e si faceva inavvertitamente coincidere la nozione di dato con la nozione di informazione. Ma anche «in informatica» il 'dato' non è identificato con l'informazione, essendo definito come «*rappresentazione* (di informazione) in un certo codice»;<sup>25</sup> dove per 'rappresentazione' si intende un insieme di segni, ovvero un insieme di «oggetti», costituiti dagli stati fisici «su cui operano gli esecutori di algoritmi», cioè gli elaboratori; e dove per 'informazione' si intendono i «significati» che sono «attribuiti ai dati», ossia all'insieme dei segni, degli oggetti, o degli stati fisici che ne costituiscono l'espressione materiale.<sup>26</sup> Certamente il dato esprime e veicola informazione, ma lo studioso è interessato primariamente all'*informazione* e utilizza il *dato* solo come un dispositivo che permette di accedervi.

In sostanza, dunque, la conoscenza della struttura del dato è necessaria per una valutazione adeguata delle operazioni che una certa forma di rappresentazione digitale permette di eseguire sul testo. La struttura di dati impiegata per ottenere la rappresentazione digitale del testo deve essere funzionale alle esigenze critiche dello studioso. Se l'esigenza è quella della semplice *lettura*, non si vede quale vantaggio aggiunga la rappresentazione digitale del testo alla sua rappresentazione convenzionale. Le prime opere a stampa impoverivano il contenuto informativo di una pagina

<sup>24</sup> A. RENEAR - E. MYLONAS - D. DURAND, *Refining our Notion of What Text Really Is: The problem of overlapping hierarchies*, in N. Ide and J. Veronis (eds.), *Text Encoding Initiative: Background and contexts*, Dordrecht, Kluwer, 1995, p. 263.

<sup>25</sup> M. LA TORRE, *Principi di informatica*, Firenze, La Nuova Italia, 1994, p. 333, corsivo aggiunto.

<sup>26</sup> *Ibid.*, p. 217.

manoscritta adeguatamente rubricata, obliterando «la volontà di usare l'articolazione visiva come mezzo di interpretazione»;<sup>27</sup> allo stesso modo, il testo digitale predisposto per la sola lettura, modificando e semplificando l'«architettura visiva»<sup>28</sup> della pagina stampata, impoverisce il contenuto informativo della sua rappresentazione convenzionale. L'elaborazione dell'informazione resta affidata all'occhio e ciò che è presentato all'occhio è decisamente più povero di ciò che gli presenta la pagina manoscritta, o la pagina di un'edizione a stampa. Se è la sola lettura, o l'accesso remoto, ciò che ne fa sorgere l'esigenza, la rappresentazione digitale cerca tutt'al più di *riprodurre* la pagina manoscritta o la pagina stampata. Non si preoccupa di utilizzare la specificità della forma digitale per elaborarne il contenuto e per aggiungere nuove funzionalità alla rappresentazione del testo. Ma la funzione di riproduzione dell'originale può essere meglio soddisfatta da un'*immagine* digitale e la forma di rappresentazione che si ottiene attraverso la codifica dei caratteri può permettere di eseguire operazioni che vanno oltre il campo della percezione visiva e integrano la sua capacità di elaborare l'informazione testuale. Ad una struttura di dati costituita da sequenze lineari di caratteri, la più semplice delle forme digitali di rappresentazione del testo, si possono quanto meno associare funzioni di ricerca automatica di stringhe corrispondenti. Ed è ovvio che in operazioni di questo tipo la macchina supera l'occhio in velocità e precisione e che l'elaborazione automatica può vantaggiosamente integrare la capacità di operare a mente sull'informazione disponibile.

La ricerca per stringhe di caratteri è un'operazione relativamente semplice, ma è su di essa che si fondano i programmi per la preparazione di concordanze. Tuttavia, per ottenere i riferimenti ai luoghi del testo da cui una certa stringa viene estratta, l'informazione testuale deve essere ulteriormente strutturata. Di solito, la tavola delle concordanze viene ottenuta operando su una struttura di dati che organizza il materiale testuale in un *database*. Il documento che costituisce la tavola è il risultato dell'esecuzione di un programma che elabora l'informazione contenuta nel *database*. In questo caso, ciò che fornisce la rappresentazione strutturale del testo è un *database*. Ed è chiaro a tutti che le strutture di dati che costituiscono un *database* sono molto più complesse della semplice struttura di dati costituita da una sequenza lineare di caratteri. Un *database* è soli-

<sup>27</sup> I. ILLICH, *Nella vigna del testo: Per una etologia della lettura*, Milano, Cortina, 1994, p. 109.

<sup>28</sup> *Ibid.*, p. 103.

tamente costituito da un'insieme di tabelle sulle quali si possono eseguire, attraverso un linguaggio di interrogazione, diverse operazioni di recupero dell'informazione; il materiale testuale deve quindi essere riorganizzato in forma non lineare per consentire operazioni più complesse sul suo contenuto. Ma è possibile identificare il testo con le strutture di dati che costituiscono le tabelle su cui opera un *database*? Di nuovo scambieremmo il dato per l'informazione che esso veicola e confonderemmo il testo con l'insieme organizzato di documenti strutturati che ne costituiscono la rappresentazione. Tali strutture di dati possono essere considerate soltanto come una forma di rappresentazione funzionale ad eseguire determinate operazioni sul testo. E l'adeguatezza di una simile rappresentazione del testo può essere valutata solo in relazione alle operazioni che permette di eseguire.

Sicché occorre interrogarsi sui «modi in cui i documenti digitali funzionano come sistemi di rappresentazione della conoscenza» per riuscire a valutare pienamente l'adeguatezza di un'edizione digitale dei materiali di studio. Lo sviluppo di «modelli» adeguati alla rappresentazione digitale del testo richiede necessariamente una conoscenza precisa delle forme in cui i documenti digitali «organizzano e strutturano le conoscenze» e comporta di conseguenza un'analisi rigorosa delle strutture di dati che li costituiscono.<sup>29</sup> Le «proprietà rilevanti dei documenti considerati come forme strutturali di rappresentazione della conoscenza» dipendono infatti dalle proprietà elementari delle loro strutture di dati.<sup>30</sup> E lo studioso che voglia utilizzare documenti digitali a fini di ricerca non può prescindere da una conoscenza esatta della natura di tale rappresentazione. Che cosa si può ottenere da una certa rappresentazione digitale della conoscenza? Le operazioni che essa rende possibili rispondono adeguatamente alle finalità della ricerca? Queste domande paiono ineludibili e stanno alla base di ogni possibile impiego di risorse digitali. Ma prima ancora di consigliarne l'impiego, esse debbono ispirarne la progettazione. E se ciò vale per i singoli documenti a maggior ragione deve valere per la progettazione e la realizzazione di intere raccolte digitali. Infatti «si può plausibilmente affermare che anche le realizzazioni più significative nel campo delle biblioteche digitali nascono da precise intuizioni circa la *struttura* dei docu-

<sup>29</sup> A. RENEAR, [*Research interests*], <URL: <http://www.stg.brown.edu/staff/affiliate/allen.html>>.

<sup>30</sup> ID., *The Digital Library Research Agenda*, «D-Lib Magazine», 3:7, 1997, <URL: <http://www.dlib.org/dlib/july97/07reear.html>>.

menti», o che le proprietà strutturali dei documenti archiviati costituiscono «in quanto tali il fondamento della funzionalità di una biblioteca digitale». <sup>31</sup>

Contrariamente a quanto comunemente si afferma, il WorldWideWeb (WWW) non può essere considerato, in quanto tale, una vera e propria biblioteca digitale. E ciò in ragione del fatto che la struttura dei documenti Web è relativamente semplice. Per rendere un documento accessibile attraverso un server WWW è sufficiente la sola codifica HTML. La codifica HTML è il solo requisito necessario per rendere un documento accessibile attraverso un server WWW. Possiamo considerare, a questo proposito, un esempio istruttivo. Una ricerca sul termine 'intensio', condotta con un comune motore di ricerca, riesce normalmente a segnalare il seguente sito Web, *Albertus-Magnus-Online: Index zur Logic*, hrsg. v. Olaf Grönemann. <sup>32</sup> Si tratta di un indice del *corpus logicum* di Alberto Magno, che intende fornire, nelle more dell'indice generale all'edizione critica completa delle opere albertine, <sup>33</sup> «un ausilio chiaramente migliore di quanto sia finora disponibile». L'indice si presenta come uno strumento prodotto «con poca spesa», ma la cui «realizzazione sarebbe impensabile senza le condizioni attualmente raggiunte dall'elaborazione elettronica dei dati». Si tratta inoltre di una «bozza di lavoro» che tuttavia «corrisponde essenzialmente, nella costruzione e nella struttura, alla progettata forma finale». L'indice è ricavato da una «concordanza KWIC» (KeyWord In Context), ossia dal risultato di una ricerca per stringhe, che è suddiviso «in singoli file HTML» secondo la lettera iniziale delle parole chiave, ovviamente per facilitare l'accesso alle singole voci attraverso una successione di rimandi ipertestuali. <sup>34</sup>

Alcune considerazioni si impongono. In primo luogo, l'indice si basa non su lemmi, ma su «forme»; dimodoché «presenta, p. es., voci distinte per *bonitas* e *bonitatem*». Inoltre «vi sono elencate tutte le forme, comprese quelle scarsamente significative come *de*, *a*, ecc.» Si tratta, evidentemente, delle conseguenze di una pura ricerca per stringhe, ottenuta su un testo non codificato. Le forme non significative non possono essere escluse e le

<sup>31</sup> *IBID.*, corsivo aggiunto.

<sup>32</sup> Version 0.1 (Arbeitsentwurf), 1998, <URL: [http://ourworld.compuserve.com/homepages/Olaf\\_Grönemann/amoidx.htm](http://ourworld.compuserve.com/homepages/Olaf_Grönemann/amoidx.htm)>.

<sup>33</sup> Cfr. ALBERTUS MAGNUS, *Opera omnia*, Münster, Aschendorff, 1951 ss.

<sup>34</sup> O. GRÖNEMANN, *Prolegomenon*, in *Albertus-Magnus-Online: Index zur Logic*, cit., <URL: [http://ourworld.compuserve.com/homepages/Olaf\\_Grönemann/amoidx.htm#prolegomenon](http://ourworld.compuserve.com/homepages/Olaf_Grönemann/amoidx.htm#prolegomenon)>

forme non possono essere organizzate per lemmi. L'ordinamento delle forme segue poi l'ordine di occorrenza nel *corpus logicum*, per cui «è allo studio un algoritmo che ordini secondo il contesto, in modo tale che formule e locuzioni come, p. es., *sylogismus dialecticus* vengano elencate nella concordanza immediatamente una sotto l'altra».<sup>35</sup> Ma lo sviluppo di un simile algoritmo richiede una strutturazione del testo che può essere ottenuta solo attraverso forme di *markup* o con l'impiego di un *database*. Si vede, ancora una volta, che i risultati prodotti dipendono dalla struttura del documento. Ma il limite maggiore di uno strumento di questo tipo dipende dalle modalità di accesso all'informazione desiderata. Come si è visto, l'indice è costituito da documenti statici o file HTML, ossia da pagine Web, e tali documenti statici sono raggiungibili solo attraverso un procedimento di 'navigazione' ipertestuale. Dall'indicazione del motore di ricerca si passa al sito, dal sito all'indice, dall'indice alla lettera iniziale, e dalla lettera iniziale, scorrendo l'elenco delle voci, all'informazione desiderata. Non ci sono altre vie per accedere direttamente al risultato. Anzi, se può capitare di aprire a caso un vocabolario alla voce desiderata, nella navigazione ipertestuale nessun passo può invece essere saltato, a meno di conoscere già, grazie ad una visita precedente, l'indirizzo specifico della pagina cercata. Ma come si può conoscere tale indirizzo prima di averlo trovato?

Sicché sono anche le limitazioni funzionali che rendono difficile considerare il WorldWideWeb come una vera e propria biblioteca digitale. Un insieme di pagine Web non può essere considerato sufficiente a formare una biblioteca digitale; in ogni caso, non più di quanto si possa pensare che un semplice cumulo di libri riesca effettivamente a costituire una biblioteca. A un'idea adeguata di biblioteca digitale risponde invece, per esempio, il progetto già in avanzata fase di realizzazione dei *Codices Electronici Ecclesiae Coloniensis* (CEEC).<sup>36</sup> Il progetto risulta infatti del tutto conforme «ai principi generali che sono emersi» da «una serie di utili discussioni»<sup>37</sup> su una recente proposta di «infrastruttura» per «servizi informativi digitali distribuiti»<sup>38</sup> e presenta tutte le caratteristiche oggi rite-

<sup>35</sup> *Ibid.*

<sup>36</sup> <URL: <http://www.ceec.uni-koeln.de/>>; cfr. M. THALLER (Hg.), *Codices Electronici Ecclesiae Coloniensis: Eine Mittelalterliche Kathedralbibliothek in digitaler Form*, Göttingen, Duehrkohp und Radicke, 2001 («Fundus - Forum für Geschichte und ihre Quellen», Beiheft 1), p. 8; pubblicato anche in forma digitale: <URL: [http://www.fundus.d-r.de/html/beiheft\\_1.html](http://www.fundus.d-r.de/html/beiheft_1.html)>.

<sup>37</sup> ARMS, *Key Concepts in the Architecture of the Digital Library*, cit.

<sup>38</sup> KAHN - WILENSKY, *A Framework for Distributed Digital Object Services*, cit.

nute essenziali per la costituzione di una «vera e propria» biblioteca digitale.<sup>39</sup> In una biblioteca digitale «l'informazione è depositata sotto forma di 'oggetti digitali'», ma «l'idea primitiva che un oggetto digitale sia solo un insieme di bit» è veramente «un'idea troppo semplice». Anche il contenuto dell'«oggetto digitale più elementare» ha «una certa struttura» e all'oggetto digitale «debbono essere associate» ulteriori informazioni, o «metadati», come la dichiarazione dei diritti di proprietà intellettuale e la specificazione dei metodi di accesso. Sicché in una biblioteca digitale «si deve distinguere chiaramente tra gli oggetti digitali quali sono creati dall'autore, gli oggetti digitali quali sono archiviati nell'infrastruttura, e gli oggetti digitali quali sono distribuiti agli utenti». Quello che l'utente riceve è «il risultato dell'esecuzione di un programma sull'oggetto depositato» e chiaramente la struttura degli oggetti e la natura dei programmi eseguibili per accedere al loro contenuto ne determinano la qualità e l'utilità per la ricerca.<sup>40</sup>

L'obiettivo dichiarato, e «intransigentemente» perseguito, del progetto di digitalizzazione dei CEEC è quello di «rendere disponibili» immagini digitali «in una risoluzione tale da rendere possibile la normale pratica di lavoro sui manoscritti».<sup>41</sup> Rispondendo in modo soddisfacente alle ordinarie esigenze di studio, la biblioteca digitale sostituisce a tutti gli effetti pratici, nella maggior parte dei casi, la consultazione diretta dei codici e con le «soluzioni realizzate» il progetto riesce a «contribuire ad una definizione di principio della forma affatto generale in cui manoscritti di grande valore possono essere resi disponibili in forma digitale».<sup>42</sup> Solo la qualità delle risorse digitali rese disponibili può quindi garantire che la biblioteca digitale possa effettivamente diventare «un luogo concreto della ricerca».<sup>43</sup> E la qualità delle risorse non può essere valutata da una «statistica dei contatti»<sup>44</sup> che si riduca al puro computo degli accessi al sito. Come può l'accesso ad una pagina 'in costruzione' risultare probante del-

<sup>39</sup> ARMS, *Key Concepts in the Architecture of the Digital Library*, cit.

<sup>40</sup> *Ibid.*

<sup>41</sup> M. THALLER, *Die Handschriftenbibliothek des Kölner Dom*, in ID. (Hg.), *Codices Electronici Ecclesiae Coloniensis*, p. 21; cfr. ID. *Digitale Archive: Technik und Methode*, in H.-H. EBELING - M. THALLER (Hg.), *Digitale Archive: Die Erschließung und Digitalisierung des Stadtarchivs Duderstadt*, Göttingen, 1999, pp. 125-62.

<sup>42</sup> ID., *Vorwort*, in *Codices Electronici Ecclesiae Coloniensis*, p. 8.

<sup>43</sup> G. SCHMITZ, *Die Bibliothek und der Bücherwurm, oder: Was kann man von digitalisierten Handschriftenbeständen erwarten?*, in *Codices Electronici Ecclesiae Coloniensis*, p. 54.

<sup>44</sup> ZORZI, *'Reti Medievali': An initiative*, cit.

la qualità di ciò che si promette di pubblicarvi? Ma anche prescindendo da possibili casi limite, «ci sono poche cose al mondo» che possono essere considerate «altrettanto irrilevanti e prive di senso per il successo di un progetto di digitalizzazione, quanto il puro e semplice numero dei contatti sul suo server». Il numero delle visite al sito può essere indice di un'esigenza, ma non dice nulla sul modo in cui tale esigenza viene soddisfatta e sulla qualità delle risorse messe a disposizione. Occorre invece definire criteri efficaci per valutare «l'uso effettivo delle risorse create». <sup>45</sup>

Nel caso dei manoscritti l'«uso effettivo» può essere garantito solo se la qualità dell'oggetto digitale che li restituisce è tale da permettere di considerarne la creazione come una vera e propria edizione della fonte. Ora, è possibile «classificare gli oggetti digitali secondo il grado in cui permettono di sostituire funzionalmente l'originale». Così, gli oggetti digitali possono essere detti 'illustrativi', quando permettono di valutare se «l'accesso diretto all'originale è utile oppure no»; 'leggibili', quando «permettono di accedere a tutta l'informazione che il creatore del documento originale intendeva comunicare», trascurando, per esempio, le aggiunte interlineari o le glosse marginali; 'paleografici', quando «permettono di accedere a tutta l'informazione che che si può ricavare a occhio nudo dall'originale»; oppure 'potenziabili', quando «la versione digitale permette di accedere ad informazione che non può essere estratta senza l'ausilio di strumenti ottici». In breve, si può affermare che «le biblioteche digitali che non offrono le risoluzioni necessarie per un lavoro professionale sono inutili». <sup>46</sup>

Ma la restituzione, anche in forma potenziabile, dell'originale non può essere giudicata una vera e propria edizione. È vero che «sempre, quando si verifica un salto di qualità tecnologico, viene ridefinito di conseguenza anche il concetto di 'pubblicazione'». <sup>47</sup> Nel Medioevo, «quando un autore aveva messo su pergamena la sua opera e l'aveva lasciata liberamente circolare perché tutti potessero leggerla e trascriverla, egli l'aveva, con i mezzi del suo tempo, 'pubblicata'». <sup>48</sup> Analogamente, dopo l'introduzione della fotografia, la stessa Direzione Centrale dei Monumenta Germaniae Historica «prese parte», nel 1905, alla pubblicazione in 'origi-

<sup>45</sup> M. THALLER, *From the Digitized to the Digital Library*, in «D-Lib Magazine», 7:2, 2001, <URL: <http://www.dlib.org/dlib/february01/thaller/02thaller.html>>, corsivi aggiunti.

<sup>46</sup> *Ibid.*

<sup>47</sup> SCHMITZ, *Die Bibliothek und der Bücherwurm*, p. 46.

<sup>48</sup> *Ibid.*, p. 45.

nale' di «un'opera di storia», segnalandola con «evidente soddisfazione»<sup>49</sup> come «una riproduzione perfettamente riuscita».<sup>50</sup> Sicché, come tempo fa la produzione fotografica di un facsimile aveva potuto, in qualche caso, essere giudicata «di grande valore» dalla comunità degli studiosi,<sup>51</sup> allo stesso modo, la possibilità di accesso alla riproduzione digitale di un manoscritto fa sì, oggi, che «i manoscritti inediti siano ubiquamente disponibili, ossia di fatto *pubblicati*»; e, «ai fini della pubblicazione, l'accessibilità ha ora preso il posto che prima spettava al manoscritto».<sup>52</sup> Ma «col fiorire dello storicismo e della critica filologica», ci si è anche resi conto che «un manoscritto non trasmette un'opera in quanto tale, ma solo una sua forma particolare»,<sup>53</sup> e alla domanda su «quale sia il senso di un'edizione di fonti, considerata dal punto di vista della sua (presente e futura) utilizzazione», pare potersi rispondere nel modo seguente:

l'edizione critica è la riproposizione (*Erschließung*) competente e scientificamente affidabile di una fonte da un punto di vista filologico e oggettivo, e [si] potre[bb]e anche aggiungere, avendo un determinato pubblico in vista; quello che si deve offrire all'utente, non può essere né troppo, né troppo poco [...]»<sup>54</sup>

Alla riproposizione della fonte deve quindi accompagnarsi

la documentazione su che cosa, dove e in quale contesto si è scritto a proposito di un certo manoscritto, dove è stato utilizzato per qualche edizione, quale posizione occupa nella gerarchia della tradizione del testo, eventualmente in quali altri luoghi si trova il testo tradito, se ne esistono trascrizioni o se l'antigrafo è conosciuto, e così via.<sup>55</sup>

Ora, una simile documentazione viene effettivamente fornita dai CEEC con la pubblicazione in rete dei materiali, sia già pubblicati a stampa, sia disponibili solo in forma elettronica, che riguardano lo studio dei singoli codici,

<sup>49</sup> ID., *Darstellung von Quellen in elektronischer Form*, <URL: [http://www.phil.uni-erlangen.de/~p1ges/reimers/schmitz/darstellung\\_frame.html](http://www.phil.uni-erlangen.de/~p1ges/reimers/schmitz/darstellung_frame.html)>; cfr. *Die Dresdner Handschrift der Chronik des Bischofs Thietmar von Merseburg*, mit Unterstützung der Generaldirektion der Kgl. Sächs. Sammlung für Kunst und Wissenschaft der König-Johann-Stiftung und der Zentralkommission der Monumenta Germaniae Historica in Faksimile herausgegeben (1905).

<sup>50</sup> Cfr. *Neues Archiv der Gesellschaft für ältere deutsche Geschichtskunde*, 31, 1906, p. 12.

<sup>51</sup> SCHMITZ, *Darstellung von Quellen in elektronischer Form*, cit.

<sup>52</sup> ID., *Die Bibliothek und der Bücherwurm*, p. 46.

<sup>53</sup> *Ibid.*

<sup>54</sup> ID., *Darstellung von Quellen in elektronischer Form*, cit.

<sup>55</sup> ID., *Die Bibliothek und der Bücherwurm*, p. 46.

come per esempio, accanto ad altri contributi o edizioni,<sup>56</sup> l'analisi di Gerhard Schmitz del codice 124,<sup>57</sup> che può essere così consultata «non più in una pubblicazione celebrativa (*Festschrift*) di qualche sorta, ma precisamente nel luogo al quale propriamente appartiene, a Colonia, accanto al manoscritto»;<sup>58</sup> o per essere più precisi, accanto alla sua riproduzione digitale.

E non è privo di importanza, al fine delle nostre considerazioni, il modo in cui l'accostamento è stato realizzato. Il lettore che voglia consultare direttamente le pagine del codice a partire dalle descrizioni ad esso associate, è rinviato immediatamente alle loro immagini digitali, visualizzabili in quattro modalità o risoluzioni diverse, a seconda delle esigenze di studio che si possono di volta in volta presentare. Nel caso dell'edizione settecentesca associata al manoscritto, sono state fornite le immagini digitali integrate da indici che permettono di rintracciare i vari testi raccolti nel codice con un rinvio all'immagine della pagina corrispondente. L'accesso e la consultazione ordinaria dei manoscritti avviene secondo diverse modalità (segnatura, contenuto, autore, titolo del testo, periodo, e così via) e i vari indici sono prodotti dinamicamente alla richiesta di accesso secondo la modalità prescelta. A ciascuna voce dell'indice sono poi associati alcuni bottoni, che permettono di visualizzare dinamicamente la documentazione, o i 'metadati' associati al manoscritto, secondo diverse forme di descrizione e di catalogazione, ciascuna delle quali rinvia direttamente alle immagini digitali del codice nelle diverse modalità di presentazione. Solo un'ispezione diretta del sito può permettere un apprezzamento adeguato delle modalità di accesso ai materiali digitali. Ma quello che qui importava mettere in evidenza è sostanzialmente la forma affatto dinamica dell'accesso all'informazione. Tutta l'informazione di volta in volta visualizzata viene estratta da un *database*, che la raccoglie e la organizza, al momento stesso della richiesta. Le strutture e il modello di dati su cui si fon-

<sup>56</sup> Cfr. Gerhard SCHMITZ, *Addenda zu Co. Köln 124: Auszugreihen und Herkunft, Literaturnachträge* (Ungedruckt); Gerhard SCHMITZ, *Initienverzeichnis zu Cod. Köln, Dizesan- un Dombibl. 124* (Ungedruckt); [Auszug aus:] Luc D'ACHERY - Louis-François-Joseph DE LA BARRE (Hgg.): *Specilegium sive collectio veterum aliquot scriptorum qui in Galliae Bibliothecis delituerant I* (Paris 1723), [S. 509-563] (Mit Initienverzeichnis), <URL: <http://www.ceec.uni-koeln.de/>>.

<sup>57</sup> Gerhard SCHMITZ, *Die Vier-Bücher-Sammlung des Cod. Köln, Dizesan- un Dombibl. 124: Zur kirchenrechtlichen Kenntnis im 10. Jh., in Ex Ipsis Rerum Documentis: Beiträge zur Mediävistik*, Festschrift für Harald Zimmermann zum 65. Geburtstag, hgg. von Klaus Herbers, Hans Henning Kortüm und Carlo Servatius (1991), S. 233-255, <URL: <http://www.ceec.uni-koeln.de/>>.

<sup>58</sup> ID., *Die Bibliothek und der Bücherwurm*, p. 54.

da il funzionamento del *database* sono ciò che determina la qualità e la funzionalità delle risorse digitali rese disponibili ai fini della consultazione e dello studio dei materiali manoscritti. Ma, di nuovo, possono le decisioni riguardanti la struttura e l'organizzazione dell'informazione all'interno del *database* che la gestisce essere totalmente delegate ai 'tecnici'? O non deve piuttosto, lo studioso che voglia effettivamente utilizzare risorse digitali a fini di studio e di ricerca, cercare di comprendere che tra le discipline ausiliarie della ricerca storica può oggi essere annoverata anche quell'*informatica umanistica*, o come altro la si voglia denominare, che nelle affermazioni più avanzate dei suoi cultori rivendica legittimamente il titolo di «scienza»?<sup>59</sup> E una vera e propria «scienza dell'informazione storica», non richiede forse una conoscenza adeguata e rigorosa delle strutture e delle forme di rappresentazione digitale delle conoscenze storiche?

Solo sulla base di una simile conoscenza l'editore di fonti storiche in forma digitale può essere in grado di svolgere adeguatamente un compito che è, e che non può essere altro che suo. Sicché «non è affatto il caso di temere, come molti fanno, che l'editore venga sostituito da un 'ingegnere del testo', capace di smontare e rimontare il testo edito senza alcuna partecipazione e discernimento». Piuttosto, «sarebbe bene» che fossero gli studiosi «a liberarsi dalla fobia del contatto con tutto ciò che viene da loro solitamente rubricato come 'tecnica'». <sup>60</sup> Infatti «c'è veramente il pericolo che la produzione di testi e di altri materiali elettronici venga marginalizzata e ignorata dagli altri studiosi del campo». E di ciò, purtroppo, paiono responsabili non tanto la diffidenza, quanto la conoscenza superficiale e generica delle tecnologie dell'informazione e la mitizzazione acritica e inconsapevole dei nuovi *media*, ora tanto alla moda. Anche l'accostamento acritico alle nuove forme di rappresentazione della conoscenza e l'incapacità di apprezzare adeguatamente la qualità delle risorse digitali, di riconoscerla e di utilizzarla propriamente a fini di studio, fanno sì che si possa anche «arrischiare l'ipotesi che sarebbe forse stato più facile familiarizzare un copista medievale con l'uso di una moderna macchina fotocopiatrice, che non accostare molti medievisti odierni all'uso del computer». <sup>61</sup>

<sup>59</sup> Cfr. M. THALLER, *Historical Information Science: Is there such a thing?*, in *Seminario discipline umanistiche e informatica: Il problema dell'integrazione* (Roma, 8 ottobre 1991), a cura di T. Orlandi, Roma, Accademia Nazionale dei Lincei, 1993 (Contributi del Centro Linceo Interdisciplinare Beniamino Segre, 87), pp. 51-86.

<sup>60</sup> SCHMITZ, *Darstellung von Quellen in elektronischer Form*, cit.

<sup>61</sup> ID., *Von Quellen und Editionen: Anmerkungen eines Mediävisten zu Problemen der Texterschließung und -darstellung*, in «Concilium medii aevi», 3, 2000, p. 57.