

# Dig *Italia*

Anno V, Numero 2 - **2010**

Rivista del digitale nei beni culturali

ICCU-ROMA



in collaborazione con ATHENA



ICCU

Istituto centrale per il catalogo unico  
delle biblioteche italiane e per le informazioni bibliografiche  
<http://www.iccu.sbn.it>

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista  
è ammessa con obbligo di citazione

# Digitalia

Rivista del digitale nei beni culturali  
ISSN 1972-6201  
Anno V, Numero 2 - Dicembre 2010

*In copertina:*

L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.c. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicicabile all'Apollo di Leochares (IV sec. a.c.)

**Direttore Fondatore**

Marco Paoli

**Direttore Responsabile**

Rosa Caffo

**Comitato di Redazione**

Anna Maria Maiorano Mandillo (coordinatore)

Lucia Basile

Simonetta Buttò

Marta Cardillo

Silvana De Capua

Eva Gilmore

Cristina Magliano

Mario Sebastiani

Vittoria Tola

**Grafica & Impaginazione**

Roberta Micchi Design

**Produzione e Stampa**

Futura Grafica S.r.l.

Via Anicio Paolino, 21

00178 Roma

**Editore**

ICCU

Istituto centrale per il catalogo unico

delle biblioteche italiane

e per le informazioni bibliografiche

Viale Castro Pretorio, 105

00185 Roma

T +39 06 49.89.484

F +39 06 49.59.302

<http://digitalia.sbn.it>

In attesa di registrazione al Tribunale di Roma



ICCU

### **Comitato Scientifico**

Osvaldo Avallone  
Pio Baldi  
Giuseppe Basile  
Giovanni Bergamin  
Armida Batori  
Simonetta Buttò  
Rossella Caffo  
Gisella Capponi  
Gabriella Contardi  
Flavia Cristiano  
Gianfranco Crupi  
Gisella De Caro  
Maurizio Fallace  
Gigliola Fioravanti  
Antonia Ida Fontana  
Paolo Galluzzi  
Daniela Grana  
Tullio Gregory  
Maria Guercio  
Mauro Guerrini  
Claudia Leoncini

Claudio Leonardi  
Cristina Magliano  
Anna Maria Maiorano Mandillo  
Maria Vittoria Marini Clarelli  
Massimo Menna  
Maurizio Messina  
Laura Moro  
Marco Paoli  
Massimo Pistacchi  
Amedeo Quondam  
Antonia Pasqua Recchia  
Don Stefano Russo  
Marco Santoro  
Mario Sebastiani  
Giuliana Sgambati  
Giovanni Solimine  
Maria Carla Sotgiu  
Laura Tallandini  
Anna Maria Tammaro  
Costantino Thanos  
Vittoria Tola  
Paul Weston

# SOMMARIO

dicembre 2010

## SAGGI

- Manoscritti nella rete** 9  
di Francesca Niutta
- OPAC & dintorni:  
essere o non essere nella rete** 29  
di Antonio Scolari
- Il paradosso della conservazione digitale:  
riflessioni sull'autenticità** 41  
di Giovanni Michetti

## PROGETTI

- ATHENA: un ponte tra i musei europei ed Europea** 57  
di Maria Teresa Natale
- Il trattamento delle immagini digitali  
alla Biblioteca estense universitaria  
di Modena (1990-2010)** 62  
di Andrea Palazzi
- Imago et umbra:  
Programma di digitalizzazione per l'Archivio storico  
della Pontificia Università Gregoriana:  
criteri, metodi e strumenti** 79  
di Andrea Bozzi, Martín Maria Morales e Marco Rufino
- La digitalizzazione dei Plutei laurenziani:  
prime considerazioni a conclusione del progetto** 100  
di Sabina Magrini
- Digitisation of Western Medieval Manuscripts  
at the British Library** 112  
di Claire Brey
- The ENRICH project:  
Towards a European digital manuscript library** 117  
di Matthew James Driscoll

**“Internet Culturale. Cataloghi e Collezioni digitali delle biblioteche italiane”. La nuova versione del portale**  
di Laura Ciancio **123**

**I progetti di digitalizzazione del Museo Bodoniano di Parma**  
di Andrea De Pasquale **135**

**Magazzini digitali: dal prototipo al servizio**  
di Giovanni Bergamin e Maurizio Messina **144**

## DOCUMENTI E DISCUSSIONI

**Delivering Content to Europeana in Practice: The ATHENA Harvesting Format LIDO**  
di Regine Stein **157**

**Final Report: Digital Libraries: Recommendations and Challenges for the Future = Relazione finale: Biblioteche digitali: raccomandazioni e sfide per il futuro**  
dell' i2010: Digital Libraries, High Level Expert Group, Copyright Subgroup **161**

## EVENTI

**4th Conference of LIBER Manuscript Librarians Group: "Meeting with manuscripts, today and tomorrow"**  
di Valentina Longo **173**

**Conferenza "CLEF 2010: Conference on Multilingual and Multimodal Information Access Evaluation"**  
di Maristella Agosti e Nicola Ferro **176**

## SEGNALAZIONI

**Antonella Agnoli, Le piazze del sapere**  
di Marta Cardillo **183**

**Roberto Raieli, Nuovi metodi di gestione dei documenti multimediali: principi e pratica del MultiMedia Information Retrieval**  
di Manuela Corbosiero **186**

**Saggi**



# Manoscritti nella rete

**Francesca Niutta**

già Biblioteca nazionale centrale di Roma

*All'inizio ci furono solo pagine miniate avulse dal contesto, tratte dai "tesori" delle biblioteche. Solo da poco più di un quinquennio hanno cominciato ad apparire nella rete riproduzioni complete di manoscritti medievali, consultabili gratuitamente. Intento del presente contributo è tracciare un panorama delle maggiori collezioni di manoscritti presenti nella rete e delle strategie che i diversi paesi perseguono al riguardo. Ma si osserva in primo luogo la difficoltà, in assenza di censimenti, a individuare il materiale nella rete, che pregiudica anche la possibilità di uno sfruttamento adeguato da parte del pubblico.*

*Un solo paese, la Svizzera, ha in corso la pubblicazione online dell'intero patrimonio manoscritto nazionale; la Spagna ha un progetto di database collettivo diretto da un'organizzazione centrale. Altrove le iniziative sono decentrate. Tra le biblioteche maggiori, la Bayerische Staatsbibliothek di Monaco e la Bibliothèque nationale de France stanno portando avanti la pubblicazione in rete di tutti i loro fondi manoscritti, mentre la British Library ha puntato sui codici greci grazie ad una sovvenzione particolare. Numerose biblioteche hanno creato database locali. In Italia ci si è concentrati sulla pubblicazione integrale di due fondi manoscritti, rappresentativi della varietà del nostro patrimonio, i Plutei della Biblioteca Medicea Laurenziana e il Fondo antico della Biblioteca del Sacro Convento di Assisi. Alcuni progetti, più complessi perché implicano un preventivo lavoro di identificazione e localizzazione dei manoscritti, mirano a sfruttare le possibilità della rete di riunire virtualmente esemplari conservati in luoghi diversi. Come quelli che riguardano le opere di un autore (Dante online) o la ricostruzione di biblioteche disperse.*

**N**egli ultimi anni Novanta, quando già la Bibliothèque nationale de France con Gallica riversava in Internet in formato immagine il *corpus* degli autori francesi del XIX secolo e oltre oceano l'American memory della Library of Congress si accresceva rapidamente con i materiali più eterogenei, scritti, visivi e sonori, l'offerta online di manoscritti medievali era limitata a qualche assaggio dei "tesori" delle biblioteche, pagine miniate avulse dal contesto, comunque per la prima volta universalmente visibili e quindi utili forse a suscitare qualche curiosità ma con effetto tantalizzante su chi il manoscritto avrebbe voluto vederlo per intero. Riproduzioni integrali di manoscritti si ebbero grazie alla Kongelige bibliotek di Copenhagen, la biblioteca nazionale danese, che cominciò nel 1997 a pubblicare in rete una serie di quelli che chiamò "facsimili digitali". Pure i manoscritti medievali, perché di questi si intende trattare in particolare, erano fra i candidati più qualificati per la digitalizzazione in quanto rispondenti a tutti i requisiti previsti, rappresentando un materiale raro (di più, ogni manoscritto è un *unicum*) e fragile,

quindi da preservare e tutelare, e libero da diritti d'autore. A distanza di anni seguirono i manoscritti di San Gallo, i Codices electronici ecclesiae Coloniensis, gli e-codices della Svizzera. E sorsero siti dedicati ai codici miniati; manoscritti e libri illustrati hanno sempre la precedenza, forse perché si ritiene che siano in grado di attrarre un pubblico di visitatori più ampio, e i numeri valgono sempre molto. I manoscritti online sono aumentati man mano e continuano ad aumentare, anche se la febbre della digitalizzazione scoppiata da poco più di un anno con Google non li ha sfiorati. Difatti i manoscritti rimangono un mondo a sé, perché richiedono precauzioni e cautele particolari nella movimentazione e nella scansione, perché le immagini non bastano, occorre un corredo complesso di informazioni sia per il contenuto che per la descrizione fisica, e forse perché si presume che abbiano un pubblico ristretto.

### **Dove, quali, quanti?**

Raccolte digitali sistematiche in via di realizzazione dei manoscritti di una biblioteca (Bayerische Staatsbibliothek, Bibliothèque nationale de France) o di un paese (e-codices della Svizzera, peraltro caso finora unico), singoli fondi digitalizzati integralmente (i Plutei laurenziani, il Fondo antico di Assisi), raccolte di manoscritti omogenei (i codici greci della British library), database di codici miniati; e inoltre siti dedicati ad un autore (Dante) o ad un'opera (*Roman de la Rose*) e progetti di ricostruzione online di biblioteche antiche smembrate (Clairvaux, Nonantola, Corvina); in più codici rubricati come tesori, o *highlights*. Questo è il ventaglio dell'offerta in Internet di manoscritti medievali e umanistici, messi a disposizione gratuitamente – come viene sempre puntualizzato – da parte di istituzioni universalmente note o di biblioteche meno conosciute. Ma le iniziative sono frammentate ed è difficile venirne a conoscenza; non esistono infatti quegli indici nazionali che erano stati auspicati agli albori della digitalizzazione (v. gli ormai remoti Principi di Lund del 2001)<sup>1</sup>; per catturare i manoscritti nella rete dobbiamo visitare un po' a caso i siti Web delle biblioteche e affidare nei motori di ricerca generali e in un paio di tentativi di censimento realizzati negli Stati Uniti che, pur lacunosi, sono di qualche aiuto. Come il Catalogue of digitized medieval manuscripts<sup>2</sup> del Center for medieval and Renaissance studies della University of California, Los Angeles, un database che fornisce links a varie centinaia di collezioni «which we expect quickly to grow to thousands», dicono i creatori. È datato 2007, ma non risulta se sia stato aggiornato successivamente. Contiene (30 ottobre 2010) 3.114 manoscritti distribuiti in 129 sedi. Consente lo scorrimento per localizzazione, segnatura, autore, titolo, lingua. Una scheda molto semplice contiene i dati essenziali del manoscritto e il link al sito che lo espone. Il maggior numero dei codici è costituito dagli e-co-

<sup>1</sup> [ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund\\_principles-en.pdf](ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund_principles-en.pdf).

<sup>2</sup> <http://manuscripts.cmrs.ucla.edu/>.

dices svizzeri e da quelli della Bayerische Staatsbibliothek; per il resto sembrano presenze in larga misura casuali; c'è un link per inviare suggerimenti su nuove collezioni da includere. Vi troviamo siti piuttosto periferici, come quello dello Árni Magnússon Institute for Icelandic studies di Reykjavík, con 51 manoscritti in norvegese antico. Per l'Italia sono elencati il Progetto Irnerio del Centro interdisciplinare in storia e filosofia del diritto e informatica giuridica dell'Università di Bologna sui codici del Collegio di Spagna<sup>3</sup>, consultazione previo contributo annuo di 60 euro e sottoscrizione di contratto di 10 pagine (è l'unico finora a pagamento); un manoscritto del Collegio di S. Isidoro a Roma<sup>4</sup>, sette codici musicali trentini del Quattrocento<sup>5</sup>, accompagnati dalla trascrizione e dalla relativa esecuzione musicale registrata; e un codice musicale di Torino, Biblioteca nazionale universitaria (J.II.9)<sup>6</sup>, con sfogliamento tridimensionale e fruscio dei fogli – artificio che si va assai diffondendo – che fa parte della collezione musicale di Internet culturale. Mancano le maggiori realizzazioni italiane. Più ampio è l'elenco curato da Robert D. Peckham della University of Tennessee at Martin (UTM)<sup>7</sup>; «made in Tennessee to bring you the world» è il suo motto. È diviso in tre sezioni: collezioni, singoli manoscritti, pagine scelte; ognuna include un centinaio di indirizzi con relativi links. Per l'Italia ignora i progetti della Biblioteca digitale italiana ma registra il catalogo aperto Malatestiano, il Dante online e la Biblioteca digitale della Biblioteca Augusta di Perugia<sup>8</sup>, che contiene 21 corali del XIII e XIV secolo della chiesa di S. Domenico.

Passeremo in rassegna alcune delle maggiori collezioni online, con l'avvertenza che altre ne esistono<sup>9</sup> e che il quadro è in continuo divenire.

## Danimarca

Pioniera nella pubblicazione in rete di facsimili digitali di manoscritti medievali e moderni (di Hans Christian Andersen e di Søren Kierkegaard per arrivare più di recente a Karen Blixen e al suo *Out of Africa*) è stata, già nel 1997, la Kongelige bibliotek di Copenhagen. Nelle due collezioni virtuali<sup>10</sup> Codices Latini Haunienses e Middle ages and Renaissance, in continuo accrescimento, sono riuniti manoscritti, ora un centinaio, appartenenti a vari fondi; gli Haunienses sono divisi in Auctores

<sup>3</sup> <http://irnerio.cirsfid.unibo.it/>.

<sup>4</sup> [http://www.tertullian.org/manuscripts\\_apologeticum/rome\\_isidore\\_1\\_29.htm](http://www.tertullian.org/manuscripts_apologeticum/rome_isidore_1_29.htm).

<sup>5</sup> <http://www1.trentinocultura.net/portal/server.pt?open=514&objID=22652&mode=2>.

<sup>6</sup> <http://www.internetculturale.it/upload/sfogliatori/franco-cip/francocip.jsp?s=6&l=en#copertina>.

<sup>7</sup> <http://www.utm.edu/vlibrary/mdmss.shtml>.

<sup>8</sup> <http://cdwdoc.demo.alchimedia.it/index1.aspx?prov=cor>. Visto il 16.11.2010.

<sup>9</sup> Progetti di numerosi altri paesi che qui non hanno potuto essere presi in considerazione si trovano illustrati nel sito del Gruppo dei bibliotecari di manoscritti di LIBER (Ligue des bibliothèques européennes de recherche) nella sezione Progress reports, 3. Digital access (<http://liber-manuscripts.kb.nl/progressreports/index.html>).

<sup>10</sup> <http://www.kb.dk/en/nb/materialer/haandskrifter/HA/e-mss/clh.html>.

antiqui, Auctores medii aevi, Auctores recentiores, Biblica, Liturgica, hagiographica, apochrypha, Ius. Le immagini sono salvabili e stampabili. La descrizione, con bibliografia aggiornata, è in inglese (talvolta è indicata anche la fascicolazione, elemento prezioso per chi non ha davanti l'originale ma consulta una riproduzione); c'è anche online il catalogo a stampa del 1926<sup>11</sup>. La sovvenzione di un mecenate privato ha consentito inoltre la digitalizzazione dei David Simonsen Manuscripts<sup>12</sup>, 194 manoscritti ebraici di varia origine e provenienza, inclusa l'Italia (31 manoscritti), presentati in un assai agevole database, con sfogliamento tridimensionale delle carte.

## Svizzera

Ma sono le biblioteche della Germania e della Svizzera che hanno messo a disposizione online nell'ultimo quinquennio il maggior numero di manoscritti medievali. La biblioteca dell'Abbazia di San Gallo con i suoi 2.100 manoscritti, di cui 400 realizzati prima dell'anno 1000, è ben nota, e ben noti sono i Codices Electronici Sangallenses (CESG)<sup>13</sup> realizzati già dal 2005 con sovvenzioni di fondazioni private, che hanno aperto la strada al successivo progetto degli e-codices, la biblioteca virtuale dei manoscritti dell'intera Svizzera in cui i sangallesi, che assommano oggi a 398, sono confluiti. Gli e-codices<sup>14</sup> svizzeri, presentati in un sito quadrilingue di ammirevole chiarezza, sono (novembre 2010) 695 di 29 diverse biblioteche. I codici della biblioteca della Fondation Martin Bodmer di Cologny, ricca di classici, faranno la felicità di tanti studiosi (sono inclusi anche due codici di Giuseppe Flavio, Cod. Bodmer 98 e 99, scritti a Nonantola nel IX secolo). Nel caso particolare si è utilizzata la descrizione di Elizabeth Pellegrin (*Manuscrits latins de la Bodmeriana*. Cologny-Genève: Fondation Martin Bodmer. 1982) di cui sono riprodotte le pagine inserendone inoltre i dati nel database catalografico dove è stata aggiunta inoltre la bibliografia aggiornata; altrimenti si sono effettuate descrizioni *ad hoc*. Nel database catalografico ad ogni pagina citata un link porta direttamente all'immagine corrispondente. Nella visualizzazione del codice si possono vedere affiancati sia recto e verso della medesima carta, sia verso e recto contigui. La legatura, con le guardie, è a parte, e solo qui figura la scala millimetrica e cromatica. Le pagine sono scaricabili e stampabili. È reciproco il passaggio immediato facsimile-descrizione.

<sup>11</sup> Ellen Jørgensen, *Catalogus codicum Latinorum medii aevi Bibliothecae Regiae Hafniensis*, Hafniae: Gyldendals, 1926.

<sup>12</sup> <http://www.kb.dk/manus/judsam/2009/sep/dsh/en>.

<sup>13</sup> <http://www.cesg.unifr.ch/it/index.htm>. Il progetto è descritto da Davide Scruzzi-Ernst Treppe, *Praktische Internet-Ausgabe und Aura des Originals. Die Stiftsbibliothek St. Gallen als Schweizer Digitalisierungs- Zentrum für mittelalterliche Handschriften*, <http://www.cesg.unifr.ch/pdf/praktische-internet-ausgabe.pdf>.

<sup>14</sup> [www.e-codices.unifr.ch](http://www.e-codices.unifr.ch).

## Germania

Annunciato nel 2001 e realizzato qualche anno più tardi dalla Biblioteca del Duomo e dall'Università di Colonia col sostegno della Deutsche Forschungsgemeinschaft, quello dei CEEC (Codices electronici ecclesiae Coloniensis)<sup>15</sup> è stato uno dei primi progetti di digitalizzazione integrale di manoscritti medievali. I CEEC sono i quasi 400 manoscritti medievali (fino al 1550) della biblioteca arcivescovile diocesana e del Duomo di Colonia, la cui creazione risale all'epoca di Carlo Magno. Il sito offre, oltre alle immagini (con scala cromatica e, non sempre, millimetrica) e alle descrizioni dei codici in triplice modalità (breve, media, lunga), un'ampia documentazione sulla propria progettazione, una ricchissima bibliografia in parte riprodotta per esteso sulla storia della biblioteca e sui singoli manoscritti, svariati cataloghi che li descrivono e materiali didattici. Si propone infatti di rivolgersi, oltre che ai destinatari naturali, anche ad un pubblico "laico", al quale è dedicato (sotto la rubrica Handschriften) un breve elenco di codici miniati nonché un manuale di introduzione ai manoscritti medievali che si può scaricare. Manca però quella scheda sintetica sulla storia della biblioteca e dei suoi fondi e sul numero e il carattere dei codici digitalizzati che il visitatore amerebbe trovare. Il database è assai articolato e consente ricerche molteplici, inclusi *incipit* ed *explicit*, sia per scorrimento che attraverso il motore di ricerca.

La Bayerische Staatsbibliothek di Monaco (BSM)<sup>16</sup>, con l'effcientissimo Münchener Digitalisierungszentrum ha realizzato una straordinariamente ricca Digital collection<sup>17</sup> (l'intero sito è bilingue; scegliamo la versione inglese) che è il contenitore di innumerevoli raccolte tematiche, repertori, opere di consultazione (ci sono anche i *Monumenta Germaniae historica*), giornali, carte geografiche, incunaboli e cinquecentine (in corso di digitalizzazione) e dal 2007 anche manoscritti. L'accrescimento procede ad un ritmo vertiginoso<sup>18</sup>. La lista sintetica delle raccolte digitalizzate si trova in Brief overview (Kurzübersicht); merita un'occhiata la base dati degli incunaboli<sup>19</sup>, le cui 9700 edizioni rappresentate nella BSM sono in via di digitalizzazione dalla primavera del 2008, con precedenza a quelle illustrate. Le immagini dei volumi sono collegate ad ampie schede catalografiche, provviste di note all'esemplare, bibliografia e link a ciascuno dei repertori citati. I manoscritti<sup>20</sup> sono divisi in una dozzina di collezioni (latini, greci, frammenti); la Project information avverte che verranno resi disponibili in Internet per gradi; la riproduzione digitale è da microfilm o da originale. Non è indicato il criterio di scel-

<sup>15</sup> <http://www.ceec.uni-koeln.de/>.

<sup>16</sup> <http://www.bsb-muenchen.de/Die-Bayerische-Staatsbibliothek.114.0.html>.

<sup>17</sup> <http://www.digitale-sammlungen.de/index.html?c=startseite&l=en&projekt>.

<sup>18</sup> 272.725 i titoli on line l'8 ottobre 2010; 397.615 un mese più tardi.

<sup>19</sup> <http://www.digitale-sammlungen.de/index.html?c=kurzauswahl&l=de&adr=inkunabeln.digitale-sammlungen.de>.

<sup>20</sup> Oltre che dalla *Digital collection* si può accedere anche da *Abteilungen*, andando poi su *Abendlandische Handschriften*.

ta dei manoscritti digitalizzati, né il numero di quelli presenti. Andiamo ai manoscritti latini, i Codices Latini Monacenses<sup>21</sup>. Lo scorrimento nella base dati è alfabetico, cronologico e per segnature; la maschera di ricerca è per autore, anno e parola-chiave del titolo (Titelstichwörter). Tentiamo una ricerca per autore; l'ovvio "Vergilius Maro, Publius", non dà risultati. Come è possibile? Allora scorriamo la lista alfabetica dei titoli; dove troviamo "Vergil, Carmina" (Clm 305) e "Virgili Aeneis" (Clm 23702). Insomma non viene dichiarato, ma l'indicizzazione per autori è ancora evidentemente *in fieri*. Nella lista cronologica diamo un'occhiata al primo manoscritto, Clm 29270 (1, un frammento di Bibbia del V e VIII secolo. Dalla descrizione sommaria si passa alle immagini e da qui al catalogo, in questo caso il recentissimo (1994) catalogo a stampa di Hermann Hauke dei frammenti latini (i cataloghi a stampa dei manoscritti sono tutti riprodotti integralmente, inclusi i più recenti). Dal catalogo l'offerta di links che portano ad un reticolo di altri links sia interni al portale che nell'intero Web è quasi infinita, e lascio allo studioso e al lettore volenteroso esplorarla. Le immagini, talvolta fornite di scala millimetrica, sono salvabili e stampabili. Per il pubblico "laico" ci sono i tesori della biblioteca in 3D (Schätze in 3D), cioè con sfogliamento tridimensionale, annunciati subito nella homepage<sup>22</sup> (mentre non è semplicissimo arrivare alle collezioni di manoscritti digitalizzati), una quindicina di codici occidentali e orientali illustrati, di vario tema e varia età, e in più la Bibbia di Gutenberg delle 42 linee (per un certo tempo c'è stato anche il *Liber chronicarum* di Schedel, scomparso all'inizio di novembre 2010).

Per conoscere le altre collezioni digitali delle biblioteche tedesche c'è Mediaevum.de. Das altgermanische Internetportal, un ricchissimo portale dedicato al medioevo tedesco<sup>23</sup>, ottimamente realizzato, che ne presenta un'ampia lista<sup>24</sup>. Per esempio c'è quella dell'Università di Heidelberg<sup>25</sup> che sta progressivamente mettendo online le collezioni di manoscritti della biblioteca, inclusi 848 codici Palatini Germanici, corredati di esauriente descrizione che include i dati codicologici. La Herzog August Bibliothek di Wolfenbüttel offre nella WDB (Wolfenbütteler Digitale Bibliothek) sia una scelta di pezzi «rari, importanti, frequentemente usati, o attualmente di particolare rilievo per la ricerca», sia una serie eclettica di progetti<sup>26</sup> miranti a fornire, oltre alle immagini digitali e alla descrizione catalografica, links alle pagine della bibliografia, e molto altro. Per la serie di manoscritti di-

<sup>21</sup> [http://www.digital-collections.de/index.html?c=kurzauswahl&l=en&adr=mdz1.bib-bvb.de/~db/ausgaben/gesamt\\_ausgabe.html?projekt=1157467155&ordnung=sig&recherche=ja](http://www.digital-collections.de/index.html?c=kurzauswahl&l=en&adr=mdz1.bib-bvb.de/~db/ausgaben/gesamt_ausgabe.html?projekt=1157467155&ordnung=sig&recherche=ja).

<sup>22</sup> <http://www.bayerische-landesbibliothek-online.de/3d>.

<sup>23</sup> <http://manuscripta.mediaevum.de>; presentazione in inglese alla pagina <http://english.mediaevum.de/>.

<sup>24</sup> <http://manuscripta.mediaevum.de/manuscr1.htm>.

<sup>25</sup> <http://www.ub.uni-heidelberg.de/helios/digi/handschriften.html>.

<sup>26</sup> <http://www.hab.de/bibliothek/wdb/index-e.htm> (versione inglese); i progetti sono elencati alla pagina <http://www.hab.de/bibliothek/wdb/projekte-e.htm>.

digitalizzati<sup>27</sup> è disponibile un elenco topografico con link alle immagini, mentre è in preparazione una banca dati indipendente con le loro descrizioni. Attira la nostra attenzione un manoscritto del gesuita Athanasius Kircher che fu ospite del Collegio Romano, il Cod. Guelf. 3.5.Aug.4°, che contiene il *Novum inventum linguarum omnium*, del 1600. Tutte le immagini sono fornite di scala millimetrica e cromatica. In Germania si devono menzionare ancora i 323 manoscritti della Sächsische Landesbibliothek – Staats – und Universitätsbibliothek (SLUB) di Dresda che fanno parte della sua Digitale Bibliothek pubblicata col supporto del Dresdner Digitalisierungszentrum (DDZ)<sup>28</sup> che ha creato un software open source ricco di molte funzioni, fra cui la possibilità di scaricare l'intero manoscritto in formato PDF. La lista dei manoscritti si può scorrere per titolo, anno, luogo di origine (ma perché non per autore?)<sup>29</sup>; c'è a parte una lista delle segnature<sup>30</sup>. Le immagini sono collegate, oltre che al database catalografico, alle pagine digitalizzate del catalogo a stampa. Merita ancora un cenno, sebbene al di fuori dei nostri confini cronologici e tematici, il Bach Digital<sup>31</sup>, frutto dello sforzo congiunto di varie istituzioni sostenuto da vari sponsor, che mira a riunire virtualmente tutti gli originali di Johann Sebastian e degli altri musicisti della famiglia sparsi in varie sedi.

## Italia

Dagli inizi del 2001 la Biblioteca digitale italiana<sup>32</sup>, nell'includere da subito i manoscritti nei suoi programmi, cominciò col promuovere la digitalizzazione dei loro inventari e dei cataloghi storici<sup>33</sup>, spesso scritti a mano, consultabili solo sul posto (in seguito molte biblioteche straniere hanno seguito l'esempio di mettere a disposizione online i cataloghi dei manoscritti, anche quelli a stampa). E poi finanzia due importanti progetti di digitalizzazione di fondi manoscritti, quello del Fondo antico del sacro convento di Assisi e quello dei Plutei laurenziani.

Il progetto assisiense, ideato dalla Società internazionale di studi francescani (SISF)<sup>34</sup> è stato il primo realizzato. Il Fondo antico del sacro convento di Assisi<sup>35</sup> è ricco di oltre 700 manoscritti, dei quali 70 miniati, che costituiscono la biblioteca, risalente al XIII secolo, della comunità francescana insediata presso la chiesa di San Francesco, finora nota a pochi. Il database contiene 260.000 immagini a 72 dpi, non stampabili, e le schede descrittive dei codici nel formato Manus. Si può scorrere la lista dei mano-

<sup>27</sup> <http://www.hab.de/bibliothek/wdb/mssdigital.htm#mss>.

<sup>28</sup> <http://digital.slub-dresden.de/dresdner-digitalisierungszentrum/?type=class%252525253DI>.

<sup>29</sup> <http://digital.slub-dresden.de/sammlungen/kollektionen/handschriften-2/nachTitel/>.

<sup>30</sup> <http://digital.slub-dresden.de/sammlungen/titelliste-nach-signaturen/>.

<sup>31</sup> <http://www.bach-digital.de>.

<sup>32</sup> <http://www.librari.beniculturali.it/genera.jsp?s=31>.

<sup>33</sup> <http://cataloghistorici.bdi.sbn.it/code/index.asp>.

<sup>34</sup> Attingo le notizie dalla presentazione del progetto nel sito della SISF, <http://sisf-assisi.it/digitalizzazione.htm>.

<sup>35</sup> <http://88.48.84.154/bbw/jsp/volumes>.

scritti per collocazione, autore, titolo, data; l'interrogazione è prevista per segnatura e per autore (non attiva quest'ultima) o per ricerca libera nel testo. Un rapido sondaggio mostra che il contenuto è più sfaccettato di quanto ci si aspetterebbe: a parte i padri della chiesa, e naturalmente in particolare S. Bonaventura da Bagnoregio, troviamo Cicerone, rappresentato da quattro manoscritti, uno dei quali (ms. 312, con la pseudociceroniana *Rhetorica ad Herennium*) presente già nell'inventario del 1381, e poi Orazio, Stazio, e ancora il *De viris illustribus* pseudopliniano.

Ma il maggior progetto italiano è quello del fondo Plutei<sup>36</sup> della Biblioteca Medicea Laurenziana di Firenze (1.500.000 euro, durata 30 mesi). Nella homepage della biblioteca<sup>37</sup> ci accolgono il Catalogo aperto e la Teca digitale, che conterrà (il completamento è prossimo) la riproduzione degli oltre 3.900 manoscritti del fondo e dei relativi cataloghi. Il progetto, realizzato col supporto scientifico della Società internazionale per lo studio del medioevo latino di Firenze, è stato da poco esaurientemente illustrato da Franca Arduini e Sabina Magrini<sup>38</sup>. Insieme ai codici sono riprodotti integralmente online i cataloghi settecenteschi che ne descrivono doviziosamente il contenuto (Bandini per i greci, latini e italiani, Biscioni per gli ebraici, Assemani per gli orientali), in base ai quali è stata fatta l'indicizzazione per segnatura, autore, titolo e data. Eventuali descrizioni moderne sono accessibili nel Catalogo aperto a cui ciascun codice è collegato e che contiene anche molte altre informazioni – la bibliografia in primo luogo – e mette a disposizione gli schedoni della consultazione con le firme in ordine cronologico degli studiosi che hanno visto il manoscritto. Il Catalogo aperto offre in più uno spazio dove gli utenti possono, previa registrazione, pubblicare i loro contributi. Tutte le immagini si possono salvare e stampare; le scale cromatica e millimetrica compaiono sulla color chart iniziale del manoscritto. Di grande utilità per lo studioso sono altre funzioni di cui il sistema è dotato; come salvare in un carrello personale, stampare, inviare via mail i risultati delle ricerche.

È d'obbligo, anche se forse superfluo, ricordare che nell'Internet culturale della Direzione generale per le biblioteche, gli istituti culturali e il diritto d'autore<sup>39</sup> rea-

<sup>36</sup> Una sintetica storia del fondo ha tracciato di recente Ida Giovanna Rao, *Il fondo manoscritto, in I manoscritti datati della Biblioteca Medicea Laurenziana di Firenze, I, I Plutei 12-34*, a cura di Teresa De Robertis, Cinzia Di Deo e Michaelangiola Marchiaro, Firenze: Sismel-Edizioni del Galluzzo, 2008, p. 3-15.

<sup>37</sup> <http://www.bml.firenze.sbn.it/>.

<sup>38</sup> Franca Arduini, *Biblioteca Medicea Laurenziana: come un'istituzione antica ha progettato il suo futuro. Digitalizzare per catalogare* in «Biblioteche oggi», vol. 28, n. 5, giugno 2010, p. 3-10; Sabina Magrini, *Il catalogo aperto della Biblioteca Medicea Laurenziana*, *ivi*, pp. 10-20; inoltre Emiliano Degl'Innocenti, *Il Progetto di digitalizzazione dei Plutei della Biblioteca Medicea Laurenziana di Firenze*, in «DigItalia», II (2007), p. 103-114, online all'indirizzo [http://digitalia.sbn.it/upload/documenti/digitalia20071\\_DEGLINNOCENTI.pdf](http://digitalia.sbn.it/upload/documenti/digitalia20071_DEGLINNOCENTI.pdf). V. ora inoltre l'articolo di Sabina Magrini, *La digitalizzazione dei Plutei laurenziani: prime considerazioni a conclusione del progetto*, in questo stesso fascicolo, p. 100-111 (n.d.r.).

<sup>39</sup> <http://www.internetculturale.it/genera.jsp?lingua=it&l=it>.

lizzato dall'ICCU, ricco di un gran numero di collezioni digitali di materiali diversi (periodici, cartografia, classici italiani inclusa la collana Scrittori d'Italia di Laterza), 2.500.000 immagini per 75.000 documenti, si trovano i fondi musicali manoscritti digitalizzati di oltre venti istituzioni<sup>40</sup> (15.000 documenti per un milione di immagini).

Si deve ad un'iniziativa indipendente, e non è una biblioteca manoscritta digitale poiché il fine primario non è, o non è solo, di mettere a disposizione codici digitalizzati, il ben noto Catalogo aperto della Biblioteca Malatestiana di Cesena<sup>41</sup> che si propone di sfruttare la possibilità che la rete offre di interazione fra biblioteca e utenti. Nato nel 2003 da un'idea di Marco Palma e Antonio Cartelli<sup>42</sup> comprende, oltre ad un database implementabile con la descrizione dei manoscritti, le loro immagini (aumenta progressivamente il numero delle riproduzioni integrali) e la bibliografia, un'apposita sezione dedicata al contesto di cui i manoscritti sono parte con materiali editi sulla biblioteca e sui suoi fondi; e un forum online, gestito dalla biblioteca, dove si possono «segnalare materiali, scambiare informazioni, pubblicare contributi inediti», nell'intento di promuovere la partecipazione all'elaborazione e la condivisione della conoscenza. Una newsletter con gli aggiornamenti è inviata periodicamente a tutti gli iscritti al forum. Preme segnalare ancora che la sostenibilità tecnologica del Catalogo aperto malatestiano è garantita dall'uso di un applicativo (Winisis) fornito gratuitamente dall'Unesco, che ne assicura il mantenimento e l'aggiornamento costante<sup>43</sup>.

## Francia

La Bibliothèque nationale de France (BnF), che per prima in Europa aveva pubblicato in Internet con Gallica un gran numero di libri digitalizzati integralmente, dal 2008 col sostegno del Centre national du livre ha avviato la Numérisation de masse (100.000 opere l'anno, 2500 documenti a settimana) che comprende una nutrita serie di programmi<sup>44</sup>. Gallica si definisce "bibliothèque encyclopédique et raisonnée" e dà accesso a documenti su ogni tipo di supporto; include, stando alle statistiche, 6338 documenti manoscritti<sup>45</sup>. Non siamo tuttavia riusciti a trovare una

<sup>40</sup> <http://www.internetculturale.it/genera.jsp?s=405&l=it>.

<sup>41</sup> <http://www.malatestiana.it/manoscritti/>.

<sup>42</sup> Marco Palma, *Il catalogo aperto dei manoscritti. Riflessioni sulle esperienze in corso*. Seminario internazionale *Zenit e Nadir II. I manoscritti dell'area del Mediterraneo: la catalogazione come base della ricerca* (Montepulciano, 6-8 luglio 2007), che si legge online: <http://dida.let.unicas.it/links/didattica/palma/testi/palmav.htm>;

<sup>43</sup> Antonio Cartelli-Andrea Daltri-Paola Errani-Marco Palma-Paolo Zanfini, *Il catalogo aperto dei manoscritti malatestiani: bilancio e prospettive*, in *Conoscere il manoscritto: esperienze, progetti, problemi. Dieci anni del progetto Codex in Toscana. Atti del convegno internazionale, Firenze 29-30 giugno 2006*, a cura di Michaelangiola Marchiaro e Stefano Zamponi, Firenze Sismel - Edizioni del Galluzzo, 2007, p. 153-163.

<sup>44</sup> [http://www.bnf.fr/fr/professionnels/selectionner\\_politique\\_programmes/s.selectionner\\_programmes\\_numertisation.html?first\\_Art=non](http://www.bnf.fr/fr/professionnels/selectionner_politique_programmes/s.selectionner_programmes_numertisation.html?first_Art=non).

<sup>45</sup> <http://gallica.bnf.fr/content?lang=fr#fonds>.

lista dei manoscritti digitalizzati, né un programma per la loro digitalizzazione. Entrando da Gallica nella base dati Manuscris<sup>46</sup> ci piace essere accolti (8 novembre 2010) da *L'esprit des lois* di Montesquieu. Possiamo sfogliare il manoscritto cliccando sulle pagine nel "mosaico", o direttamente visualizzandolo a schermo intero; stampare le pagine, scaricarle in uno "spazio personale", inviarle per e-mail. Ma non riusciamo da qui ad arrivare alle collezioni di manoscritti. Per raggiungerli il percorso è tortuoso: si deve entrare in Collections et départements, andare a Département des manuscrits, poi a BnF archives et manuscrits<sup>47</sup>, dove è spiegato che il catalogo dei manoscritti è in via di costituzione, e dove si può usare il formulario di ricerca o accedere alle collezioni, e anche raggiungere i cataloghi del dipartimento, quasi tutti digitalizzati in modo immagine. Cliccando sul nome della collezione appaiono le segnature; cliccando sulla segnatura può apparire una descrizione sommaria oppure una scheda completa redatta appositamente, o niente del tutto; la descrizione è a volte accompagnata dalla riproduzione integrale del manoscritto (es. Lat. 1090, Lat. 8041, Lat. 8125); oppure ci può essere la riproduzione senza descrizione catalografica (es. NAL 255). Insomma ci troviamo nel pieno di un cantiere di lavoro, l'embrione di una digitalizzazione di massa anche dei manoscritti.

La BnF ha in corso anche una serie di progetti in collaborazione con istituzioni straniere<sup>48</sup>. Il Rose project, finanziato dalla Andrew W. Mellon Foundation (\$779.000), frutto della collaborazione fra BnF e The Johns Hopkins Sheridan Libraries di Baltimora, porterà alla digitalizzazione di 150 manoscritti del *Roman de la Rose* appartenenti alla stessa BnF e ad altre biblioteche francesi. Consentirà di studiare insieme ornamentazione e poesia ma intende anche, mediante una documentazione di carattere didattico, rivolgersi a un pubblico ampio.

Nel grandioso progetto Europeana regia<sup>49</sup> «which aims to create a European corpus of digitised, mostly illuminated manuscripts», la Bibliothèque nationale de France è affiancata da un nutrito gruppo di partners: Bibliothèque Royale de Belgique, Bayerische Staatsbibliothek, Herzog August Bibliothek Wolfenbüttel, Biblioteca Històrica-Universitat de València (non l'Italia). Europeana regia si articola in tre diverse raccolte: Bibliotheca Carolina, che conterrà 425 manoscritti considerati i capolavori delle principali abbazie dell'impero carolingio (VIII e IX secolo); la biblioteca del re Carlo V (170 manoscritti); e, quel che ci riguarda più da vicino, la biblioteca dei re d'Aragona di Napoli (282 manoscritti) che come ben sappiamo è divisa fra tante biblioteche diverse in tutto il mondo.

Numero complessivo delle immagini: 307.000. Ciascuna delle biblioteche parteci-

<sup>46</sup> <http://gallica.bnf.fr/editors?type=manuscripts>.

<sup>47</sup> <http://archivesetmanuscrits.bnf.fr/pages/index.html>.

<sup>48</sup> Sono stati illustrati da Thierry Delcourt nella 4th Conference of LIBER Manuscript Librarians Group "Meeting with manuscripts, today and tomorrow", Roma, Biblioteca Nazionale Centrale, 26-28 maggio 2010, <http://www.bnrcrm.librari.beniculturali.it/index.php?it/390/thierry-delcourt>.

<sup>49</sup> <http://version1.europeana.eu/web/guest/details-europeanaregia/>.

panti immagazzinerà le proprie e le metterà a disposizione sul proprio sito; saranno poi raccolte nel portale di Europeana; è allo studio uno schema descrittivo multilingue. Il budget totale del progetto, finanziato al 50% dalla Comunità Europea, è di 3.400.000 euro; durata prevista: 30 mesi a partire da gennaio 2010.

In Francia sono tante le biblioteche, grandi e piccole, che stanno mettendo online il loro patrimonio di manoscritti medievali provenienti prevalentemente dalle confische delle biblioteche religiose della Rivoluzione francese. Citiamo la Bibliothèque de Reims<sup>50</sup>, 3.000 manoscritti di cui 860 medievali inclusi gli oltre 50 del IX secolo. Il nucleo più importante è quello del Capitolo della Cattedrale, legato nel IX secolo al nome dell'arcivescovo Incmaro. Col sostegno dell'Institut de recherche et d'histoire des textes, del Ministero della cultura e della regione Champagne-Ardenne ha digitalizzato e messo in linea cento manoscritti; in allestimento il database. E la Bibliothèque municipale de Lyon<sup>51</sup>, che conserva e ha digitalizzato, grazie ad una sovvenzione della MRT (Mission de la recherche et de la technologie) del Ministère de la culture et de la communication, 55 manoscritti merovingi e carolingi di quelli, circa 600, raccolti nel IX secolo dal diacono Florus nella biblioteca episcopale e oggi divisi fra varie biblioteche europee. Il sito è ammirevole e fornisce tutte (o quasi) le informazioni che si vorrebbero. Dalla lista topografica<sup>52</sup> si ha subito un'idea del contenuto e dell'età dei codici. Che contengono, come è naturale, in prevalenza testi patristici; ma vi scopriamo anche un *Timeo* di Platone nella traduzione di Calcidio, del IX secolo (ms. 324), e un frammento di Terenzio ai fogli 91-97 del ms. 788 dell'VIII secolo. Dalla lista si può andare direttamente alle immagini, stampabili, anche selezionandone un particolare. La maschera di ricerca è abbinata all'indice dei manoscritti<sup>53</sup>, che presenta a sinistra una descrizione breve, da cui si può andare alla visualizzazione delle immagini, o richiamare la descrizione dettagliata che comparirà sulla destra. Nella descrizione dettagliata, cliccando sul numero della pagina che indica il contenuto si va direttamente all'immagine relativa, come negli e-codices svizzeri. La ricerca può avvenire per collocazione, autore, titolo, data; le liste relative, cliccando, compaiono in una finestra. Quattro manoscritti (mss. 425, 431, 484, 597) possono essere sfogliati in 3D. Con fruscio della pergamena.

## Spagna

In Spagna c'è la Biblioteca virtual del patrimonio bibliográfico<sup>54</sup>, un progetto cooperativo del Ministerio de cultura e delle Comunidades autónomas, che mira alla diffusio-

<sup>50</sup> <http://www.bm-reims.fr/webcontent/viewer/viewer.asp?INSTANCE=exploitation&EXTERNALID=WBCTDOC%5F131&SYNCMENU=EN%5FCHANTIER>.

<sup>51</sup> [http://www.bm-lyon.fr/trouver/basesdedonnees/base\\_manuscrit.html](http://www.bm-lyon.fr/trouver/basesdedonnees/base_manuscrit.html).

<sup>52</sup> [http://www.bm-lyon.fr/trouver/basesdedonnees/base\\_manuscrit-liste.html](http://www.bm-lyon.fr/trouver/basesdedonnees/base_manuscrit-liste.html).

<sup>53</sup> <http://florus.bm-lyon.fr/index.php>.

<sup>54</sup> <http://bvpb.mcu.es/es/estaticos/contenido.cmd?pagina=estaticos/presentacion>.

ne mediante facsimili digitali delle collezioni di manoscritti e libri a stampa (e altro) del patrimonio storico spagnolo. La maschera di ricerca è unica per tutto il materiale digitalizzato, ma si può selezionare il tipo di documento desiderato, “material textual manuscrito” nel nostro caso. Troviamo 1530 manoscritti in tutto (uno dell’VIII secolo, quattro del IX, 22 del X). Possiamo selezionare lingua, secolo, biblioteca, materia/soggetto del contenuto. Le opzioni di scorrimento nella lista che appare sono per importanza, titolo, autore, data. Dalla scheda si va alle immagini, che si possono esportare e stampare, anche parzialmente (in PDF); le notizie catalografiche si possono scaricare, stampare, inviare per e-mail, salvare in uno schedario personale. La Biblioteca nacional di Madrid cura «con el apoyo de: Telefonica», come si legge in ogni pagina, un’altra collezione digitale online, la Biblioteca digital hispánica, ricca di oltre 20.000 documenti, divisi in sezioni corrispondenti alle varie discipline<sup>55</sup>. I manoscritti sono classificati insieme a “libros notables”, “bibliofilia” ma è previsto un filtro per selezionarli. Quindici quelli ora presenti, a cui è da aggiungere la musica manoscritta (circa 750 documenti).

### Gran Bretagna

Una delle prime raccolte di facsimili digitali era stata realizzata in Gran Bretagna con gli Early manuscripts at Oxford University<sup>56</sup>; conta ora un’ottantina di esemplari di sette biblioteche oxoniensi tutti realizzati sugli originali. Ma è preclusa purtroppo allo studioso che non abbia elevate capacità informatiche poiché la visualizzazione richiede un’apposita configurazione della macchina. La National library of Wales di Aberystwyth nel suo Digital mirror, che mira principalmente alla valorizzazione della cultura del Galles, include anche ma non solo manoscritti in lingua galles<sup>57</sup>. La British library ha da poco (27 settembre 2010) messo in linea la riproduzione delle pagine scritte di 287 manoscritti greci, 250 della Additional manuscripts collection, 37 della Harley<sup>58</sup>; qualcosa di leggermente diverso dunque dai “facsimili digitali”, perché mancano le legature, che in qualche caso sono accessibili a parte. Il progetto, finanziato dalla Stavros Niarchos foundation, è stato curato da un team ragguardevole di conservatori e studiosi. Come viene spiegato, la scelta è caduta sugli Additional perché offrivano un buon panorama dei diversi tipi di manoscritti greci e includevano alcuni capolavori, e inoltre perché erano fra i meno accessibili essendo privi di una catalogazione moderna. La scelta degli Harley invece ha tenuto conto delle risposte al quesito «Which manuscripts should we digitise?» pubblicato sull’apposito blog Medieval and earlier manuscripts. Il browsing, ascendente o discendente, è per segnatura, autore, titolo (a scelta in greco o in latino), scriba; la ricerca avanzata si può effettuare per parole chiave, data, collocazione, titolo, autore/scri-

<sup>55</sup> <http://bdh.bne.es/bnearch/>.

<sup>56</sup> <http://image.ox.ac.uk/>.

<sup>57</sup> <http://www.llgc.org.uk/index.php?id=digitalmirror-manuscripts>.

<sup>58</sup> <http://www.bl.uk/manuscripts/About.aspx>.

ba, provenienza/acquisizione, bibliografia. La scheda catalografica, appositamente realizzata per gli Additional, nella descrizione fisica include il tipo di rigatura e per i cartacei la filigrana, ma non purtroppo, in genere, la fascicolazione; riporta la storia del codice e la bibliografia. Nella descrizione del contenuto dei miscellanei di fianco ad ogni testo c'è la relativa immagine *thumbnail* da cui si va direttamente alla visualizzazione della pagina digitalizzata. Il programma di visualizzazione è estremamente semplice e funzionale. Ogni immagine, corredata di scala millimetrica, reca segnatura, data, titolo e autore, numero della carta; si possono affiancare verso e recto successivo ("open book") e recto e verso dello stesso foglio ("folio"). Nella stampa compare solo l'immagine *thumbnail*.

### Stati Uniti

La Library of Congress, riccamente dotata di archivi di autori contemporanei anche di notevole importanza (Sigmund Freud, Anna Arendt), è povera di manoscritti medievali. Ma merita qualche parola l'American memory con cui aveva dato il via dagli anni Novanta alla presentazione massiccia online dei materiali più diversi, dalle lettere di Abraham Lincoln alla musica tradizionale degli Indiani di Omaha, sempre corredata di ampie introduzioni e commenti. Le Digital collections dell'American memory, raccolte sotto rubriche che vanno da "Advertising" a "Women's history"<sup>59</sup>, hanno una esclusiva funzione didattica e di rado offrono documenti nella loro integrità; per lo più contengono solo alcune riproduzioni (a bassa risoluzione, difficilmente leggibili) a supporto dei testi, spesso ampi saggi, che presentano queste "primary sources", dandone regolarmente la trascrizione<sup>60</sup>. Anche le altre biblioteche statunitensi mostrano in genere scarsa propensione – ci sono anche eccezioni – a mettere in linea riproduzioni integrali, limitandosi per lo più a esibire qualche pagina esemplificativa.

Il Digital scriptorium<sup>61</sup>, decano dei database di manoscritti medievali nordamericani, presenta una scelta di immagini, in accordo col suo carattere. Non è infatti una biblioteca digitale manoscritta, sebbene venga comunemente considerato tale; si autodefinisce «online visual catalogue of medieval and Renaissance manuscripts» e contiene per ciascun manoscritto una serie di immagini che valgono a integrarne la descrizione catalografica<sup>62</sup>. Creato nel 1997 dalle risorse congiunte delle univer-

<sup>59</sup> <http://memory.loc.gov/ammem/index.html>.

<sup>60</sup> All'importanza delle *primary sources* è dedicata una pagina: <http://www.loc.gov/teachers/usingprimarysources/>. Sono espressamente diretti agli insegnanti programmi per la conoscenza delle fonti («Bringing the power of primary sources into the classroom»), mirabilmente realizzati: <http://www.loc.gov/teachers/>.

<sup>61</sup> <http://www.scriptorium.columbia.edu/>.

<sup>62</sup> Consuelo Dustchke, "Digital scriptorium": *ten years old*, in *Conoscere il manoscritto* cit., p. 189-205. Più di recente La Dutschke è intervenuta sulla sostenibilità del Digital scriptorium: *Digital scriptorium: ten years young and working on survival*, in «Storicamente» 4 (2008) che si può leggere online: [http://www.storicamente.org/02\\_tecnostoria/filologia\\_digitale/dutschke.html](http://www.storicamente.org/02_tecnostoria/filologia_digitale/dutschke.html).

sità di Berkeley e Columbia, è in continuo accrescimento, rimanendo fedele ai suoi principi: digitalizzare, per limitare la spesa, alcune immagini di *tutti* i manoscritti di una biblioteca, piuttosto che procedere alla digitalizzazione completa di un numero ristretto di esemplari. Le schede sono brevi poiché, per lo stesso principio, la scelta è di una catalogazione estensiva anziché intensiva. Le informazioni ricavabili dalle immagini (per esempio i colori delle miniature) non sono incluse nella descrizione: la funzione delle parole è assunta dalle immagini stesse. La ricerca nel database è per localizzazione, segnatura, autore, titolo, copista, miniatore. Oggi (ottobre 2010) le istituzioni aderenti sono trenta, i manoscritti più di 5.000, le immagini 27.000. Due links nella homepage del Digital scriptorium mandano a due censimenti di manoscritti statunitensi, quello dei manoscritti petrarcheschi (pubblicato sotto l'égida del Comitato per le celebrazioni del VII centenario della nascita di Francesco Petrarca con finanziamento del Ministero per i beni e le attività culturali, Direzione generale per i beni librari e gli istituti culturali)<sup>63</sup> e quello dei greci (una quarantina attualmente; molti altri annunciati).

Invece il *Roman de la Rose* della Johns Hopkins University<sup>64</sup>, che potrebbe definirsi un portale monografico, è ispirato ad un principio opposto: raccogliere tutte le immagini di tutti i manoscritti del testo cui è dedicato. Il progetto risale al 1996, quando alcuni docenti (Stephen G. Nichols, James M. Beall) della Johns Hopkins University cominciarono a procurarsi delle copie digitali di manoscritti del *Roman de la Rose* a scopo didattico, e si è progressivamente allargato coinvolgendo studiosi e istituzioni diverse. L'obiettivo è di rendere accessibili in rete i circa 300 manoscritti noti, presenti del resto già in buon numero nel database, dislocati in Europa e in America, fornendoli anche di ampi sussidi didattici, inclusa la trascrizione del testo. La visualizzazione include lo sfogliamento tridimensionale.

Manoscritti digitalizzati integralmente presenta la Houghton Library dell'Università di Harvard, che partecipa anche al Digital scriptorium, e ha una interessante raccolta di codici in latino, greco e nei volgari dell'Europa occidentale<sup>65</sup>. Nell'elenco dei latini<sup>66</sup> (solo topografico) scegliamo il Ms. Lat. 188, datato 1466, con le *Facetiae* di Poggio Bracciolini. Tutte le immagini sono accompagnate dalla scala millimetrica e cromatica. Si possono, dieci alla volta, convertire in un file PDF temporaneo stampabile. Il record catalogafico si raggiunge cliccando sul nome dell'autore.

Invece la Morgan library & museum, che detiene una delle più rinomate raccolte statunitensi di manoscritti medievali<sup>67</sup>, ricca in particolare di miniati, offre solo una immagine per ciascun esemplare accompagnata da record bibliografico; un link porta alla descrizione dettagliata offerta dal catalogo vero e proprio digitalizzato,

<sup>63</sup> <http://www.franciscus.unifi.it>.

<sup>64</sup> <http://romandelarose.org/#home>.

<sup>65</sup> [http://hcl.harvard.edu/libraries/houghton/collections/early\\_manuscripts/index.cfm](http://hcl.harvard.edu/libraries/houghton/collections/early_manuscripts/index.cfm).

<sup>66</sup> [http://hcl.harvard.edu/libraries/houghton/collections/early\\_manuscripts/bibliographies/Lat.cfm](http://hcl.harvard.edu/libraries/houghton/collections/early_manuscripts/bibliographies/Lat.cfm).

<sup>67</sup> <http://www.themorgan.org/collections/collectionsMedRen.asp>.

il Corsair catalog, con link a sua volta alla bibliografia più recente.

Sulla scia della Library of Congress si muove la miriade di biblioteche statunitensi, universitarie e pubbliche, con piccole raccolte di manoscritti, che insieme ai materiali digitalizzati offrono ampi supporti didattici.

La Free library of Philadelphia presenta la riproduzione integrale dei codici<sup>68</sup> accompagnata da una guida<sup>69</sup> e un glossario e fornisce anche una piccola bibliografia ragionata per l'introduzione alla conoscenza dei manoscritti; le immagini recano un commento semplice e chiaro, rivolto a chi non ha familiarità con essi. Nella maschera di ricerca proviamo col solito "Vergilius", senza successo. Perché si deve scrivere "Virgil"! E così troviamo 4 manoscritti. Quando poi procediamo alla ricerca per secoli, fra il 900 e il 925 troviamo 23 frammenti di un codice greco in minuscola bouletée, Lewis E 251, riprodotti integralmente. Insomma l'intento divulgativo che il sito persegue non esclude l'offerta di materiali rivolti allo studioso.

Diverso lo stile adottato dalla New York Public library. Fra le infinite raccolte tematiche digitali ne ha una dedicata ai manoscritti dal IX al XVI secolo<sup>70</sup>, con riproduzione di più di 2000 pagine, corredata di una lista dei soggetti da cui si va alle immagini digitalizzate; la descrizione del manoscritto è sommaria e appare solo se si va a cliccare. Impossibile però arrivare agli autori e ai titoli dei testi, perché non c'è un indice. Peccato. Perché scorrendo le immagini si possono scoprire cose interessanti; per esempio il famoso codice Aulendorf della cronaca del Concilio di Costanza di Ulrich von Richental (Spencer collection Ms. 32), acquerellato, con Giovanni Hus sul rogo e le sue ceneri gettate nel Reno. L'ingrandimento disponibile non consente la lettura del testo.

### Raccolte virtuali

La maggioranza delle basi dati online mira a presentare i manoscritti che una biblioteca possiede. Ma ci sono anche progetti che mirano a utilizzare la possibilità della rete di riunire virtualmente esemplari conservati in luoghi diversi, più complessi perché implicano un preventivo lavoro di identificazione e localizzazione dei manoscritti. Come quelli che riguardano le opere di un autore, o di un testo, o la ricostruzione online di collezioni antiche smembrate fra varie sedi. Dei portali monografici del Bach digital tedesco e del *Roman de la Rose* franco-americano si è detto. Il portale Dante online<sup>71</sup>, progetto dell'ente Cassa di risparmio di Firenze realizzato con la consulenza scientifica della Società dantesca italiana, oltre a presentare la vita e l'opera di Dante, aspira a raccogliere le riproduzioni di tutti i manoscritti delle sue opere (sono 827 quelli censiti, 33 quelli ora visionabili)<sup>72</sup>, for-

<sup>68</sup> <http://libwww.freelibrary.org/medievalman/>.

<sup>69</sup> <http://libwww.freelibrary.org/medievalman/guide.cfm>.

<sup>70</sup> [http://digitalgallery.nypl.org/nypldigital/explore/?col\\_id=173](http://digitalgallery.nypl.org/nypldigital/explore/?col_id=173).

<sup>71</sup> <http://www.danteonline.it>.

<sup>72</sup> [http://www.danteonline.it/italiano/codici\\_indice.htm](http://www.danteonline.it/italiano/codici_indice.htm).

nendo anche la trascrizione dei testi. Peccato quella filigrana invasiva tipica delle digitalizzazioni italiane che attraversando tutta la pagina deturpa le immagini e non agevola la lettura.

Quanto alla ricostruzione di biblioteche antiche smembrate, è partito da poco il progetto riguardante la biblioteca napoletana dei re d'Aragona. La Bibliotheca Corviniana digitalis<sup>73</sup> si propone la ricostruzione virtuale della biblioteca di re Mattia Corvino divisa fra una dozzina e più di paesi, della quale l'Ungheria conserva 53 manoscritti, tutti online nel sito della biblioteca nazionale ungherese che ha promosso il progetto.

Vedrà la luce entro il 2011 la ricostruzione virtuale della biblioteca dell'abbazia di Clairvaux, la maggiore biblioteca medievale in Francia, ricca nel 1472 di 1.790 manoscritti; ne sono rimasti 1.115; il maggior numero è conservato dalla Médiathèque de l'Agglomération Troyenne<sup>74</sup>, che ne curerà la realizzazione in collaborazione con l'Institut de recherche et d'histoire des textes e le istituzioni francesi e straniere fra le quali sono divisi oggi i manoscritti. E invece non si sa quando potrà essere realizzato il progetto di ricostruzione online della biblioteca del monastero di Nonantola, uno dei più illustri dell'Italia medievale, sede anche di un importante *scriptorium*. Il nucleo maggiore di 45 codici compresi fra il VI e il XII secolo è conservato dalla Biblioteca Nazionale Centrale di Roma, che ne ha effettuato la digitalizzazione integrale e ha elaborato dal 2004 il progetto di Biblioteca nonantolana virtuale<sup>75</sup>; gli altri sono stati identificati e localizzati da una serie di studiosi in una dozzina di istituzioni italiane ed europee. La banca dati con i 45 manoscritti digitalizzati è per il momento consultabile nella Sala manoscritti e rari della Biblioteca.

### Manoscritti miniati

Svariati siti europei sono dedicati ai codici miniati. In genere riportano solo qualche pagina illustrata ma includono una scheda con la descrizione completa del manoscritto, e quindi risultano utili anche a chi non si interessi specificamente di miniatura. Sono tutti in via di arricchimento progressivo. Le immagini, spesso a bassa risoluzione, sono scaricabili e stampabili. Il sistema di classificazione e indicizzazione dei soggetti è il più vario, avendo ciascuna istituzione elaborato il proprio – con l'eccezione della Koninklijke bibliotheek dell'Aia che usa una classificazione codificata.

Tre ne presenta la sola Francia nel portale Manuscrits enluminés des bibliothèques de France<sup>76</sup>: Mandragore, la base dati di manoscritti miniati della BnF e della

<sup>73</sup> <http://www.corvina.oszk.hu>.

<sup>74</sup> [http://patrimoine.agglo-troyes.fr./CDA/portal.aspx?PAGE=/bmtroyes/projets/clairvaux2.html&INSTANCE=exploitation&PORTAL\\_ID=troyes\\_content\\_iframe.xml](http://patrimoine.agglo-troyes.fr./CDA/portal.aspx?PAGE=/bmtroyes/projets/clairvaux2.html&INSTANCE=exploitation&PORTAL_ID=troyes_content_iframe.xml).

<sup>75</sup> <http://www.bncrm.librari.beniculturali.it/index.php?it/175/biblioteca-nonantolana-virtuale>.

<sup>76</sup> <http://www.manuscritsenlumines.fr/>.

Bibliothèque de l' Arsenal, Liber Floridus e Enluminures. Mandragore<sup>77</sup> comprende (novembre 2010) più di 170.000 notizie relative a manoscritti che vanno dall'Egitto dei faraoni all'età contemporanea, indicizzate con un vocabolario di 18.000 descrittori (in francese) e ordinate secondo una classificazione tematica da Arte a Zoologia, passando per Bibbia, Storia, Letteratura, nonché Parapsicologia, Occultismo, Demonologia, con infinite sottoclassi, accompagnate da immagini digitalizzate. La scheda del manoscritto è sommaria. Due maschere consentono la ricerca attraverso la descrizione del manoscritto (per segnatura, autore, titolo, paese, luogo di origine, miniatore, data; tutti i campi sono forniti di indici), o iconografica. La base dati Liber Floridus<sup>78</sup> «a pour ambition de proposer la consultation à un large public de l'ensemble des enluminures des manuscrits médiévaux conservés dans les bibliothèques de l'enseignement supérieur». Contiene ora (novembre 2010) circa 31.000 immagini di 1.600 manoscritti delle biblioteche Mazarine e Sainte Geneviève; l'indicizzazione, straordinariamente ricca (prevede anche una voce "Codicologie", sotto la quale sono classificati i "signes codicologiques", come manicule e segni di paragrafo), è in corso. Enluminures<sup>79</sup> presenta 80.000 immagini di più di 4.000 manoscritti medievali di un centinaio di biblioteche municipali francesi; aspira anch'essa a rivolgersi ad un pubblico ampio, cui è dedicata la bella pagina Qu'est-ce qu'un manuscrit enluminé? Nelle Visites virtuelles offre perfino alcuni filmati dedicati alla realizzazione di manoscritti<sup>80</sup>. La maschera di ricerca è particolarmente "amichevole"; la scheda descrittiva del manoscritto è estremamente sommaria.

La British library ha realizzato a partire dal 2003 col concorso di vari enti, incluso "Il Circolo" Italian cultural association, il Catalogue of illuminated manuscripts<sup>81</sup>. Si possono scorrere gli indici per luoghi d'origine, scritture, scribi, artisti; le opzioni di ricerca includono il formato e la legatura (ma non è spiegato come indicarli). Le immagini sono collegate a schede complete del manoscritto; la visualizzazione è ottima e le immagini sono stampabili. C'è un glossario illustrato dei termini tecnici<sup>82</sup> (ma manca un elenco dei descrittori); si possono seguire dei tour didattici attraverso i manoscritti miniati della biblioteca. Ha tutt'altro carattere e finalità Images online<sup>83</sup>, una specie di grande catalogo di immagini di libri, manoscritti, carte geografiche diviso in sezioni tematiche e continuamente aggiornato; offre la licenza a pagamento per il loro uso a editori, giornali e riviste, agenzie di pubblicità, televisioni; lo scopo primario è commerciale, ma il sito costituisce ugualmente

<sup>77</sup> <http://mandragore.bnf.fr/html/accueil.html>.

<sup>78</sup> <http://liberfloridus.cines.fr/textes/cines.html>.

<sup>79</sup> <http://www.enluminures.culture.fr/documentation/enlumine/fr/index3.html>.

<sup>80</sup> <http://www.interbibly.fr/virtuelles/trhc/index.html>.

<sup>81</sup> <http://www.bl.uk/catalogues/illuminatedmanuscripts/welcome.htm>.

<sup>82</sup> <http://www.bl.uk/catalogues/illuminatedmanuscripts/glossary.asp>.

<sup>83</sup> <http://www.imagesonline.bl.uk/>.

una fonte di conoscenza del patrimonio della British library. Le immagini *thumbnail* appaiono in ordine casuale ma, trovata quella che fa al caso nostro, attraverso le parole chiave associate si può partire per una navigazione dagli esiti imprevedibili che può essere assai fruttuosa.

Il sito olandese Medieval illuminated manuscripts<sup>84</sup> contiene 11.000 miniature di manoscritti della Koninklijke Bibliotheek dell'Aia che lo gestisce e del Museum Meermanno-Westreenianum, classificate secondo l'Iconclass system<sup>85</sup>; la lista dei soggetti è in quattro lingue, italiano incluso; l'introduzione è di una chiarezza rara, perché «in agreement with the ambitions of a national library, publishing sources on the Internet must aim at a professional as well as a lay audience»<sup>86</sup>. La maschera di ricerca, assai articolata ma di agevole interpretazione, prevede anche il formato, espresso in millimetri, le misure dello specchio di scrittura e il numero delle linee. Ogni pagina offre la possibilità di dialogare con la biblioteca tramite e-mail per comunicare correzioni o aggiunte da apportare o per ordinare copie. Il sito include un percorso guidato, mirabilmente realizzato, attraverso le immagini dei capolavori (Highlights) rivolto «a chi non abbia un interesse professionale per la miniatura medievale», ricco di links col database delle miniature. L'Olanda appare scarsamente propensa a presentare online interi manoscritti, ma vale la pena di ricordare che soccorre in altri modi il ricercatore. Il portale Medieval manuscripts in Dutch collections<sup>87</sup> comprende, oltre all'utilissimo database con la descrizione dei manoscritti latini fino al 1550 delle collezioni olandesi, incluse quelle di archivi e musei, ausili bibliografici e didattici (con un atlante paleografico) e il catalogo dei manoscritti datati delle biblioteche olandesi.

In Italia una base dati di miniature *in progress*, Colori online, circa 7.000 immagini di 271 manoscritti, è offerta dalla Biblioteca Riccardiana di Firenze<sup>88</sup> che presenta anche, nel Parnaso di carta, alcuni esemplari particolarmente rappresentativi in facsimile digitale con sfogliamento tridimensionale e un sorprendente accompagnamento musicale. In Colori online dall'indice delle segnature si passa all'elenco delle carte miniate; c'è una scheda catalografica sommaria e un repertorio iconografico. Il programma di visualizzazione offre in più, rispetto agli altri, una lente di ingrandimento, utilissima per i particolari (e per le glosse).

## Per chi?

La pubblicazione dei manoscritti nella rete mette potenzialmente a disposizione di chiunque un patrimonio finora nascosto e sconosciuto ai più; ma sono gli studiosi,

<sup>84</sup> <http://www.kb.nl/manuscripts/>.

<sup>85</sup> È liberamente disponibile all'indirizzo <http://www.iconclass.org/>.

<sup>86</sup> <http://www.kb.nl/manuscripts/information>.

<sup>87</sup> <http://www.mmdc.nl/static/site/>.

<sup>88</sup> <http://www.riccardiana.firenze.sbn.it/main.php?Lang=IT>; cfr. Giovanna Lazzi, "Colori on line": il prototipo per una banca dati di immagini, in *Conoscere il manoscritto* cit., p. 165-170.

gli specialisti, a trarne particolarmente beneficio. Sono loro i primi destinatari delle collezioni online delle grandi e piccole biblioteche europee (della Danimarca, della Germania, della Svizzera, della Francia, dell'Italia, della Gran Bretagna, della Spagna) che abbiamo passato in rassegna. Anche se ben sappiamo che la miglior riproduzione digitale non può sostituire la visione diretta dell'originale.

È stato osservato<sup>89</sup> che i fruitori dei manoscritti online coincidono col pubblico delle sale manoscritti delle biblioteche che, si sa, sono in numero esiguo. Ma non per questo i manoscritti e le loro riproduzioni online sono sottoutilizzati. Se sono pochi i frequentatori delle sale manoscritti anche il numero complessivo dei manoscritti è relativamente esiguo (benché incalcolato) e non lontanamente comparabile con quello dei libri a stampa – oggetto ora di estesi e discussi progetti di digitalizzazione –; i manoscritti costituiscono quantitativamente una minoranza per lo più infinitesima nel patrimonio complessivo di una biblioteca, ma sicuramente il loro livello di consultazione è elevato, o forse percentualmente più elevato – non conosco calcoli al riguardo – della media totale dei volumi conservati.

La pubblicazione nella rete costituisce comunque una grande occasione di divulgazione della conoscenza; e questa è certo una via da seguire, purché se ne abbiano le risorse. Sono in particolare i siti specializzati in codici miniati, i più attraenti, che si prefiggono di rivolgersi anche ad un pubblico ampio. Così come sono diretti anche ai non professionisti quei progetti che offrono insieme alle immagini dei manoscritti la trascrizione dei testi, come il Dante online e il *Roman de la Rose*. Le biblioteche statunitensi sono in genere quelle maggiormente inclini alla divulgazione e utilizzano i manoscritti a fini didattici; però sembrano restie a metterne a disposizione degli studiosi copie integrali online.

## Domani

È annunciata la digitalizzazione degli 80.000 manoscritti della Biblioteca Vaticana, veramente «un'impresa grandiosa, a profitto della cultura e in particolare della tutela e conservazione dei beni affidati alla Biblioteca Apostolica»<sup>90</sup>. Intanto si può prevedere, nell'ambito di un aumento massiccio di documenti online incoraggiato dai governi e dall'Unione europea, un incremento anche delle collezioni manoscritte. Il Programme national de numérisation del Ministère de la culture et de la communication francese ha appena lanciato un *Appel à pro-*

<sup>89</sup> Marco Palma, *Which audiences for manuscripts?* Relazione tenuta alla 4th Conference of LIBER Manuscript Librarians Group "Meeting with manuscripts, today and tomorrow", Roma, Biblioteca Nazionale Centrale, 26-28 maggio 2010; si legge on line: <http://www.bncrm.librari.beniculturali.it/index.php?it/388/marco-palma>.

<sup>90</sup> Mons. Cesare Pasini, prefetto della Biblioteca apostolica Vaticana, Newsletter 5/2010 del 24.3.2010, [http://www.vaticanlibrary.va/home.php?pag=newsletter\\_art\\_00086](http://www.vaticanlibrary.va/home.php?pag=newsletter_art_00086). Da notare che il progetto si avvale per la conversioni delle immagini di un formato non proprietario, il formato FITS elaborato dalla Nasa, da cui è garantito quindi il costante aggiornamento.

*jets de numérisation 2011*<sup>91</sup> che include specificamente la digitalizzazione di manoscritti medievali. Il portale di Internet culturale potrebbe arricchirsi presto di nuove collezioni, poiché la Direzione generale per le biblioteche ha da poco rivolto alle biblioteche pubbliche statali l'invito a inviare dati e informazioni su progetti di digitalizzazione realizzati o in allestimento<sup>92</sup>; e per i manoscritti ce ne sono sicuramente in attesa nei cassetti.

Quando tutto ciò si avvererà, e la Bibliothèque nationale de France, la Bayerische Staatsbibliothek nonché la Spagna avranno completato la digitalizzazione delle loro collezioni con relativi database, saremo vicini alla *Bibliotheca manuscripta universalis* di Ezio Ornato<sup>93</sup>. Ma la prima condizione per la più ampia utilizzazione delle risorse in rete è l'informazione su quello che vi si trova; confidiamo che qualche rivista specializzata si assuma l'onere di una rubrica apposita con segnalazioni e aggiornamenti sul materiale manoscritto che viene progressivamente pubblicato online. A meno che non si riesca prima a interfacciare tutte le basi dati di manoscritti, come è nel progetto Manuscriptorium<sup>94</sup>, rendendo inutili indici e censimenti.

*In the beginning, there were only a few illuminated pages taken from the "treasures" of various libraries. It has only been little more than five years since the first complete reproductions of medieval manuscripts started making their appearance on the Internet, freely available. The present article wishes to provide an overview of the various manuscript collections that are now available online, and of the strategies followed by various countries on the matter. However, one of the first obstacles in doing so, is the lack of listings locating what is in fact available – and this hinders the possibility in itself for users to access these online collections.*

*There is only one country, Switzerland, currently working on the online publication of its entire national manuscript heritage. Whereas Spain is working on a project for a collective database managed by a central organization. Other countries have decided to decentralise their actions. For what concerns major libraries, both Bayerische Staatsbibliothek in Munich and the Bibliothèque Nationale de France are in the process of publishing their entire manuscript funds on the internet, while the British Library is focusing on Greek Codes thanks to a special subvention. Many libraries have also created local databases. Italy has focused its efforts on the entire publication of two manuscript funds which well represent the variety of our heritage – the Pluteus Fund of the Biblioteca Mediceo Laurenziana and the Ancient Fund of the Biblioteca del Sacro Convento di Assisi. Finally, there are a few more complex projects which require a preliminary effort to identify and locate the manuscripts, and are aimed at taking maximum advantage of the internet by virtually collecting items preserved in different places – as in the case of one author collections (Dante Online) or of the virtual rebuilding of lost libraries.*

<sup>91</sup> [http://www.culture.gouv.fr/culture/mrt/numerisation/fr/actualit/documents/appe\\_l\\_numerisation2011.doc](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/actualit/documents/appe_l_numerisation2011.doc).

<sup>92</sup> <http://www.internetculturale.it/generaNews.jsp?s=14&id=865&l=it>.

<sup>93</sup> Ezio Ornato, *Bibliotheca manuscripta universalis. Digitalizzazione e catalografia: un viaggio nel regno di Utopia?*, «Gazette du livre médiéval», n. 48, printemps 2006, p. 1-13, anche online <http://www.palaeographia.org/glm/art/utopia1.htm>

<sup>94</sup> <http://www.manuscriptorium.com/index.php>

# OPAC & dintorni: essere o non essere nella rete\*

**Antonio Scolari**

*Università degli studi di Pavia*

*Gli OPAC stanno evolvendo verso una nuova generazione di interfacce di ricerca sia dietro la spinta dei produttori di sistemi di automazione sia per la pressione delle biblioteche. A fronte però della disponibilità in rete di collezioni digitali sempre più numerose e ricche ci si può chiedere quali dovrebbero essere le loro linee di evoluzione, in particolare rispetto agli utenti che ci si propone di raggiungere. Per raggiungere utenti remoti nella rete è oramai indispensabile partecipare a cataloghi di ambizione internazionale o mondiali; infatti solo in presenza di una massa critica di dati davvero significativa e col supporto di forti investimenti si può traguardare un obiettivo che la concorrenza dei motori di ricerca generalisti rende difficilmente raggiungibile. Tuttavia nel ripensare gli OPAC non si debbono neppure perdere di vista le esigenze degli utenti locali della biblioteca, che si rivolgono al catalogo per utilizzare la collezione specifica – fisica o digitale che sia – rappresentata dal catalogo. L'evoluzione degli OPAC rappresenta quindi una sfida importante per le biblioteche, di cui occorre cogliere le opportunità, ma nel contempo riconoscere con chiarezza gli obiettivi.*

La disponibilità di numerose collezioni digitali a testo pieno in rete ha un impatto non piccolo sulle biblioteche, non solo in termini di organizzazione e fruizione delle raccolte, ma anche sulla presenza delle biblioteche nella rete. In particolare sembra via via diminuire il peso e l'importanza degli OPAC (Online Public Access Catalogue) nella rete, cioè l'interesse per utenti "remoti" ad accedere ai cataloghi delle biblioteche e viceversa viene da chiedersi se gli OPAC attuali rispondono in modo adeguato alle esigenze degli utenti "locali", quelli che ancora vengono in biblioteca o che comunque sono interessati proprio alla collezione specifica della biblioteca.

Nel seguito si useranno in modo alquanto improprio gli aggettivi "remoto" e "locale" a proposito degli utenti di un OPAC: infatti si riferiscono questi termini agli utenti non tanto dal punto di vista della vicinanza o della lontananza fisica e geografica alla biblioteca, aspetto che assume un peso molto relativo a fronte della ubiquità della rete, ma piuttosto dal punto di vista degli interessi specifici degli utenti. Così potrà essere "locale" (o se si vuole "non troppo remoto") un utente fisicamente lontanissi-

\* Questo saggio riprende l'intervento presentato al Seminario della Associazione Itale, *Nuovi modelli per nuovi scenari: i sistemi bibliotecari italiani di fronte alle trasformazioni della società della conoscenza*, Certosa di Pontignano (Siena), 19-20 aprile 2010.

mo, che però è interessato alla collezione, fisica o virtuale che sia, della biblioteca; per contro lo studente di un ateneo o il cittadino che non ha uno specifico interesse per la collezione della biblioteca pubblica o universitaria che gli è fisicamente vicina, diventa a tutti gli effetti un “utente remoto”. Sembra questa una distinzione necessaria e conseguente alla dematerializzazione che la rete introduce in modo evidente e consistente nelle pratiche di fruizione dei contenuti della rete stessa.

L'evoluzione del modo di circolazione e utilizzo della informazione nello spazio del Web induce a una riflessione sul possibile ruolo e sulla rilevanza degli OPAC, quali strumenti per il reperimento delle informazioni nella rete.

È stato giustamente considerato che:

«OPACs still have a necessary role in preserving access to library collections, maintaining a consistent and authoritative form of bibliographic control and providing a targeted information environment for specific client groups»<sup>1</sup>.

Tuttavia in un recente intervento critico, ma anche stimolante, sul ruolo degli OPAC, si osserva che:

«The World-Wide Web has become the people's encyclopedia of choice. Google and other Web search engines give people a good start, and, in fact, with Wikipedia links in hand, it gives them a running start, for building on their bare-bones, basic knowledge of a topic. The Web also satisfies people's voracious appetites for full texts [...]. Instead of strolling in the library stacks to find a book, people want to stay put in their homes and offices and retrieve full texts with a click of a button»<sup>2</sup>.

Due gli aspetti che vengono enfatizzati da Markey; un primo è riferito alla sempre minore disponibilità degli utenti a seguire l'iter tradizionale per l'accesso alle informazioni in biblioteca: nel momento in cui per molte attività della vita quotidiana non è più necessario recarsi fisicamente in uffici pubblici, banche o negozi, per rispondere a una serie di bisogni informativi, non pare necessario recarsi fisicamente neppure in biblioteca. Un secondo aspetto riguarda la percezione che gli utilizzatori della rete hanno dei punti naturali di accesso per la ricerca dell'informazione. In questo secondo caso il ruolo centrale è giocato dalla quantità di informazione a cui si ritiene di potere accedere. Infatti nello spazio informativo della rete un servizio è percepito solo se raggiunge una sufficiente “massa critica”, in termini dimensionali, per essere visto: ad esempio, la sola Wikipedia nella versione italiana è intorno alle 730.000 voci, ma sono 3.250.000 le voci di quella in inglese, e sono ben oltre sette i milioni delle voci complessive in tutte le lingue. Numeri imponenti se

<sup>1</sup> David Wells, *What is a library OPAC?*, «The Electronic Library», 25 (2007), n. 4, p. 386, doi:10.1108/02640470710779790.

<sup>2</sup> Karen Markey, *The online library catalog: paradise lost and paradise regained?*, «D-Lib Magazine», 12 (2007), n. 1-2, <http://www.dlib.org/dlib/january07/markey/01markey.html>.

confrontati con i dati correnti nei nostri OPAC. Si può aggiungere, ed è questo un ulteriore punto assai importante per il prossimo futuro, che finora il tipo di approccio degli utilizzatori agli strumenti di ricerca nella rete è stato determinato da servizi quali i motori di ricerca, innanzi a tutti Google, wiki, Web sociale, ma – ora – i cosiddetti *mobiles*, cioè l'insieme di *devices* (telefoni portatili dalle funzioni avanzate, *tablet computer* come l'ipad, lettori di e-books) e programmi loro dedicati, stanno causando un rapido e significativo cambiamento nella organizzazione dei servizi di rete e nel modo di interagire con essa.

Viene da chiedersi allora se uno dei pensieri che negli ultimi anni hanno guidato l'esposizione degli OPAC nella rete, cioè la consapevolezza che il valore bibliografico dei cataloghi esposti in rete trascende la raccolta delle biblioteche di cui rappresentano la collezione, non sia stato almeno in parte fuorviante. Ben poche biblioteche, o forse nessuna, e pochi consorzi dispongono di una massa critica tale di dati bibliografici da renderli davvero "visibili" in rete. Inoltre anche sul versante dell'approccio verso gli utilizzatori, malgrado i non pochi tentativi, gli OPAC per loro natura (qualcuno potrebbe aggiungere "per fortuna") restano molto lontani dagli strumenti appena citati, per non dire di come nelle tecnologie "*mobiles*" gli OPAC stentino a trovare una loro vera collocazione. Credo che questo derivi dalle difficoltà di diventare strumenti diffusi al di fuori di cerchie abbastanza ristrette di utilizzatori<sup>3</sup>.

Non si vuole con questo negare che il valore di strumento bibliografico attribuito ai cataloghi, agli OPAC che hanno contribuito, tra l'altro, a una maggiore apertura e diffusione dei cataloghi delle biblioteche, non abbia la sua importanza, tuttavia è legittimo domandarsi fino a che punto si debbano inseguire le evoluzioni tecnologiche col rischio di rendere meno funzionale l'utilizzo dell'OPAC all'utente locale, volendo privilegiare gli utenti remoti della rete, la cui esistenza in quanto utenti del catalogo resta però alquanto ipotetica.

Porsi questo interrogativo sembra, a dire il vero, ozioso e fuori luogo, nel momento in cui produttori commerciali di ILS e di sistemi e servizi per le biblioteche e anche comunità di sviluppo di prodotti *open source* propongono sistemi integrati di ricerca, i cosiddetti portali verticali, o *discovery interfaces*, nei quali l'OPAC evolve verso un punto informativo esteso, in grado di offrire l'accesso alle più svariate risorse di una biblioteca o di un consorzio, in qualche caso non solo strettamente bibliografiche<sup>4</sup>. Si tratta di prodotti che fanno evolvere il tradizionale concetto di

<sup>3</sup> Dal punto di vista puramente tecnologico, oramai molti cataloghi sono accessibili tramite *smartphone*, anche se non senza qualche problema, date le profonde differenze degli apparecchi mobili sul mercato. Si veda in proposito Samuel Liston, *OPACs and the mobile revolution*, «Computers in Libraries», 29 (2009), n. 5, p. 6-11, 42-47.

<sup>4</sup> Ad esempio in *Summa*, sviluppato dalla State and University Library of Denmark, è prevista la possibilità di creare "schede" per esperti di specifiche discipline, schede che essendo catalogate anche a livello semantico vengono visualizzate effettuando ricerche per soggetto. Informazioni su *Summa* all'indirizzo: <http://www.statsbiblioteket.dk/summa>.

meta catalogo o meta motore di ricerca, mutuando dai motori di ricerca generalisti, quali Yahoo o Google, la raccolta in indici locali di dati bibliografici provenienti da cataloghi, basi dati, repository, metadati di collezioni digitali. Tali portali verticali utilizzano motori di ricerca sviluppati *ad hoc* e in grado di elaborare sia informazione strutturata che non strutturata, di eseguire il *relevance ranking* dei risultati, di determinare la correlazione semantica tra concetti e consentire la navigazione a faccette. Infine integrano funzionalità più o meno spinte, derivate dal Web sociale<sup>5</sup>.

La proposta emergente del mercato dei sistemi di automazione per biblioteche è quindi quella di costruire archivi di dati sufficientemente ampi e ricchi per potere raggiungere una massa critica tale da renderli rilevanti agli occhi degli utenti remoti, nel tentativo di limitare, anche mutuando alcune delle caratteristiche del Web 2.0, la fuga sempre più massiccia degli utenti dai cataloghi e dai servizi in rete delle biblioteche.

Una ulteriore e recentissima evoluzione di questi prodotti vede la proposta di quelli che si potrebbero definire “macro-depositi” remoti di indici a cui attingere da parte di biblioteche o consorzi, a livello geografico, se non planetario<sup>6</sup>. La proposta è supportata da motivazioni più evidenti, quale ad esempio la difficoltà e i costi gestionali per il mantenimento di simili prodotti a livello locale, o anche più sottili, quali ad esempio la possibilità per i produttori di effettuare analisi d’uso e di accesso, anche sofisticate, nell’ipotesi di proporre in un futuro non lontano nuove tipologie di indici bibliometrici per la valutazione della ricerca e di entrare in un mercato sempre più interessante e appetibile<sup>7</sup>. Anche in questo caso va sottolineato il ruolo che gioca la massa critica dei dati per “esserci” nella rete e quindi ottenere una visibilità da parte degli utenti remoti dei servizi di biblioteca.

Un limite ben noto che presentano i nostri cataloghi nei confronti degli utenti remoti è di essere solo raccolte di metadati, mentre oggi siamo abituati a raggiungere subito i contenuti, magari non proprio quelli che cercavamo, ma almeno simili o surrogati dei quali ci accontentiamo perché molto spesso rispondono in modo passabilmente sufficiente alle nostre esigenze. Il numero di documenti a testo pieno

<sup>5</sup> I più noti prodotti commerciali di questo tipo sono *AquaBrowser Library* di Medialab Solutions, *Encore* di Innovative Interfaces, *Primo* di ExLibris e *Summon* di SerialSolutions.

<sup>6</sup> Va in questa direzione ad esempio il recente lancio da parte di ExLibris di *Primo Central*, ma in questi mesi sembra volere entrare in questa fascia di prodotti anche Elsevier – tradizionale produttore di contenuti – con la sua nuova interfaccia *SciVerse*, che per ora si presenta come un “hub” di raccordo tra diversi prodotti dell’editore, ma che non nasconde la sua ambizione di proporsi come interfaccia di ricerca aperta anche ad altri editori, con l’aggiunta di collegamenti a funzionalità di analisi e valutazione della ricerca (<http://info.sciverse.com/>).

<sup>7</sup> Su questa tematica si veda Johan Bollen – Herbert Van de Sompel, *An architecture for the aggregation and analysis of scholarly usage data*, in: *Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries*, Chapel Hill (NC): ACM, 2006, p. 298-307, doi:10.1145/1141753.1141821.

raggiungibili dai nostri OPAC infatti è ancora modestissimo e per lo più limitato a settori speciali della documentazione, quali i periodici elettronici per le biblioteche universitarie e di ricerca o libri fuori diritti per le biblioteche pubbliche. Né probabilmente, nel nostro paese, il numero dei *full-text* negli OPAC è destinato a crescere in modo esponenziale, poiché in ogni caso gli OPAC restano soprattutto rappresentativi della collezione fisica posseduta. Tuttavia è da considerare la possibilità che nel giro di pochi anni il mercato degli e-books, favorito dalle tecnologie *mobiles*, si estenda in modo esponenziale anche in Italia, come sta succedendo soprattutto negli Stati Uniti da un paio d'anni a questa parte, e quindi finisca per cambiare più rapidamente di quanto oggi non pensiamo i rapporti di rappresentazione tra carta e bit negli OPAC delle biblioteche.

Per queste ragioni e per i limiti intrinseci dei cataloghi, rispetto a molte delle risorse in rete, se si vuole puntare all'esposizione diffusa in rete dei metadati dei cataloghi, avendo come mira il raggiungimento degli utenti remoti delle biblioteche, parrebbe logico, e si direbbe ineluttabile, puntare davvero a un numero molto ridotto di cataloghi mondiali (di "world cat[alogues]"), magari in grado di essere propulsivi delle innovazioni della rete, e non solo di esserne all'inseguimento. Ma soprattutto è auspicabile che la loro massa critica in termini di metadati informativi cumulati li renda davvero percepibili e visibili nella rete.

Da questo punto di vista ci si può chiedere se Google Books non possa diventare o non sia già l'OPAC della rete, il vero "world cat", dal momento che Google Books ha acquisito nel tempo alcuni degli aspetti tipici di un OPAC.

Innanzitutto verrebbe da pensare che da un punto di vista numerico Google Books sia già il catalogo mondiale, certo – come minimo – con il forte svantaggio di non rappresentare una collezione, ma piuttosto una sorta di coacervo, di *bric-a-brac* di collezioni, in crescita continua, infinitamente più rapida e tumultuosa di quella di qualsiasi collezione fisica, ma fondamentalmente disordinata e casuale<sup>8</sup>. Tuttavia una qualche considerazione va fatta a proposito dei numeri del "catalogo" Google Books. Anche per questo aspetto – come per molti degli aspetti tecnologici e amministrativi – Google è assai OPACo: in un atto giurato (*affidavit*) presentato da Google nel febbraio 2010 viene dichiarata una spesa di «hundreds of millions of dollars researching, developing, patenting and implementing cutting edge digital scanning technology», denari si direbbe forse non spesi benissimo, visti i risultati spesso modesti del riconoscimento di caratteri

<sup>8</sup> Si veda la messa a punto delle numerose problematiche di Google Books in Alberto Petrucciani, *La bancarella planetaria e la biblioteca digitale: il punto di vista della ricerca e una possibile agenda per l'Italia*, «DigItalia» V (2010), n. 1, p. 9-32, [http://digitalia.sbn.it/upload/documenti/DIGIT%201-2010\\_riv.pdf?l=it](http://digitalia.sbn.it/upload/documenti/DIGIT%201-2010_riv.pdf?l=it).

<sup>9</sup> Si vedano gli esempi citati, alcuni esilaranti, in Petrucciani, *La bancarella planetaria* cit., p. 16-18. Una riprova evidente e macroscopica la si può avere provando a scaricare su un lettore di e-books libri in formato e-pub da Internet Archive: poiché la gran parte delle digitalizzazioni provengono da Google, la bassa qualità dell'OCR rende di fatto inutilizzabile questa funzionalità, pure in teoria

nei libri digitalizzati da Google<sup>9</sup>; i libri digitalizzati dichiarati erano allora «more than twelve million», diventati nel frattempo oltre quindici milioni<sup>10</sup>; a inizio 2010 erano 49 le biblioteche da cui Google aveva ricevuto metadati e 21 i database commerciali da cui acquisisce metadati, per una spesa annua dichiarata di 2.500.000 dollari; infine negli anni Google ha raccolto e analizzato 3,27 miliardi di registrazioni bibliografiche, identificando oltre 174 milioni di record unici, che dovrebbero concorrere – si immagina – a comporre l'ideale biblioteca "totale"<sup>11</sup>. I numeri, anche se imponenti, mostrano che, in particolare per quanto concerne il numero dei libri digitalizzati, Google Books è ancora ben lontano dall'essere quel repository totale, la biblioteca universale che spesso siamo indotti a credere che sia; l'osservazione è ancora più calzante se si tiene conto dell'alto numero di duplicazioni, in cui ci si imbatte effettuando ricerche su Google Books. Certamente rispetto ai cataloghi delle biblioteche Google Books ha l'essenziale vantaggio di offrire al consultatore remoto non solo metadati, per altro – come ben si sa – di dubbia qualità e di non sempre alta coerenza, ma di potere sfruttare i testi pieni delle pubblicazioni digitalizzate, anche quando non vengono resi gratuitamente disponibili al pubblico, e quindi con l'indubbio valore aggiunto di potere sfruttare il contenuto completo, offrendo viste di quei contenuti non sempre immediatamente banali tali da stupire anche utenti non certo sprovveduti. Lorcan Dempsey, ad esempio, commentava di recente sul proprio blog la scoperta della digitalizzazione di un suo vecchio libro in Google Books (figura 1), del quale i contenuti vengono visualizzati in modo non convenzionale.

Si potrebbero definire in sostanza delle funzionalità, per dirla con una terminologia di moda, da «OPAC arricchito». Allo stesso ambito appartengono l'offerta di copertine, indici e visualizzazione parziale dei contenuti, anche per i libri ancora disponibili sul mercato o appena pubblicati, che Google Books offre grazie agli accordi stipulati con molti editori di svariati paesi.

Infine tra le funzioni di Google Books ritroviamo anche quella, più tipica degli OPAC, dell'indirizzamento dell'utente a una copia fisica a lui più vicina, disponibile sia presso librerie che presso biblioteche, raggiungibili queste ultime tramite il link creato verso il catalogo WorldCat della Online Computer Library Center (OCLC).

interessante e destinata ad assumere sempre maggiore importanza con l'aumentare della disponibilità di lettori di e-books. La controprova immediata la si può avere scaricando libri, sempre nel formato e-pub da Gutenberg Project: in questo caso la buona qualità di partenza dei testi (sempre rivisti manualmente) consente nella maggior parte dei casi una lettura pressoché perfetta anche su un lettore e-books.

<sup>10</sup> Questo il dato dichiarato da James Crawford (engineering director, Google Books) in un blog del 14 ottobre 2010 <http://booksearch.blogspot.com/2010/10/on-future-of-books.html>.

<sup>11</sup> I dati e la riproduzione dell'*affidavit* in Norman Oder, *Google Book search by the numbers*, «Library Journal», 12 February 2010, <http://www.libraryjournal.com/article/CA6718929.html>.

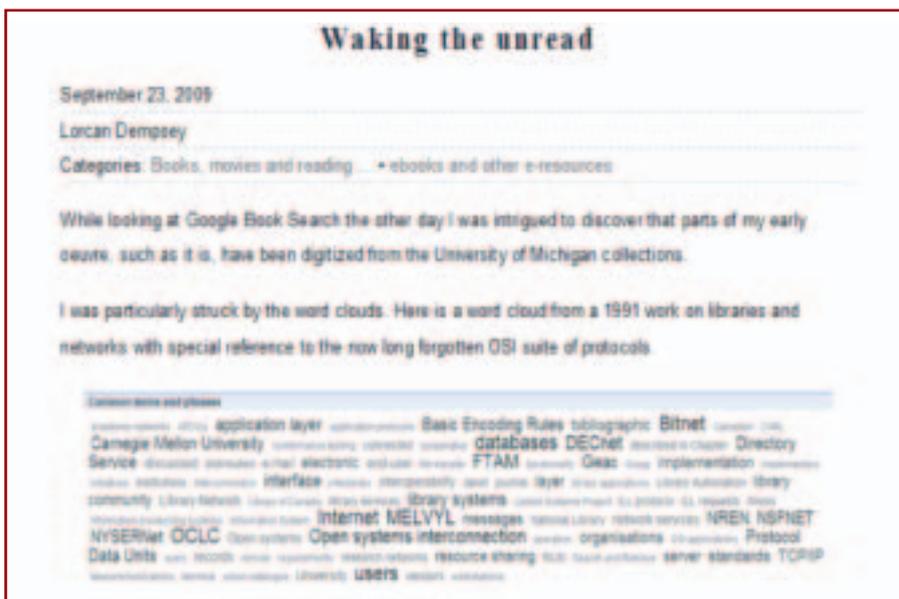


Figura 1. Dal Blog di Lorcan Dempsey

Da questo punto di vista la scelta effettuata da OCLC, oramai da alcuni anni, è stata indubbiamente significativa: avere collegato WorldCat a Google Books lo ha reso anche il principale catalogo generalista della rete e uno strumento primario di indirizzamento alle copie fisiche conservate nelle biblioteche che partecipano a quel catalogo. Inoltre all'inizio di quest'anno è stata annunciata una iniziativa complementare di OCLC: verranno caricati in WorldCat registrazioni da Google Books e da Hathi Trust<sup>12</sup> per consentire dal catalogo un accesso diretto ai dati digitali.

Oggi WorldCat conta circa 203 milioni di record bibliografici che rappresentano oltre un miliardo e mezzo di copie, SBN conta oltre 10 milioni di record per 43 milioni di localizzazioni. È evidente il peso che assume per utenti remoti WorldCat. Di conseguenza il risultato di una ricerca, rappresentato in fig. 2, per una monografia della quale l'Indice nazionale SBN conosce una settantina di localizzazioni, manifesta una indubbia assenza delle biblioteche italiane da una delle risorse più utilizzate in rete per il reperimento di libri delle biblioteche.

<sup>12</sup> Hathi è il repository che raccoglie i volumi digitalizzati, per la gran parte nell'ambito dei progetti Google, di alcune delle principali biblioteche universitarie della California, dell'Ohio e di altri stati americani. Hathi prova così ad offrire, agli utenti delle biblioteche che partecipano al progetto, una ulteriore strada, alternativa a Google e a Internet Archive, per fruire dei dati digitalizzati nel corso degli ultimi anni, offrendo tra l'altro metadati di qualità indubbiamente superiore, perché frutto di rielaborazione di dati catalografici, rispetto a quelli offerti dalle altre iniziative.



Figura 2. Esito di una ricerca in WorldCat

Il recente annuncio della collaborazione tra Ministero per i beni e le attività culturali e Google per la digitalizzazione massiva di pubblicazioni delle Biblioteche nazionali centrali induce a credere che ci si muoverà verso questo tipo di soluzioni, e in effetti un collegamento oltre a quello con la piattaforma di Google, anche con un qualche strumento diffuso delle dimensioni di un WorldCat o simile avrebbe una sua logica per la diffusione della informazione digitalizzata e consentirebbe di rimediare, almeno in parte, alla scarsa visibilità in rete dei contenuti digitali del nostro paese.

Fin qui alcune osservazioni sulle problematiche che si pongono agli OPAC per raggiungere un'utenza remota diffusa e la constatazione che fatalmente siano altri, rispetto agli OPAC, gli strumenti di ricerca che si rivolgono con successo a questo tipo di utenti. Proprio a fronte di questi aspetti è naturale pensare che sia giunto il momento di rivalutare la dimensione eminentemente "locale" dell'OPAC o se si vuole ribadire il valore di catalogo della biblioteca e quindi di pensarlo davvero a misura di utente non troppo remoto di una precisa collezione. Rispondendo, in questo modo, anche a quelle che restano comunque le richieste di maggior momento degli utenti che si rivolgono al catalogo della biblioteca: «Dove trovo il libro che mi interessa?», «Posso avere una versione in rete del documento?», ma anche «Posso cercare per soggetti più accurati?»<sup>13</sup>. È evidente la necessità di rinnovare

<sup>13</sup> Questi i risultati che emergono dal report *Online catalogs: what users and librarians want: an OCLC report*, Dublin (Ohio): OCLC, 2009, <http://www.oclc.org/reports/onlinecatalogs/default.htm>.

strumenti sviluppati in epoca oramai non vicina e soprattutto pensati come versioni meccanizzate del catalogo a schede, in particolare per quanto concerne la presentazione dei risultati; tuttavia non sempre le scelte, che paiono andare per la maggiore nei nuovi OPAC, rispondono a queste domande poste dagli utenti<sup>14</sup>. Vediamo qualche minimo esempio.

All'inseguimento di Google oramai molti OPAC offrono come prima scelta la ricerca libera per parole in un solo campo. Se si può dubitare che questo approccio sia davvero utile per un utente che viene in biblioteca o che consulta il catalogo della biblioteca e spesso ha in mente un titolo, un autore, un argomento, è però di fondamentale importanza sottolineare che questo tipo di ricerca offerta dagli OPAC non utilizza la metodologia post-booleana di Google e di altri motori di ricerca, ma utilizza ancora i vecchi e oramai obsoleti meccanismi booleani che governano l'accesso ai nostri cataloghi, come pure a molte basi dati, portali e altri strumenti bibliografici<sup>15</sup>. Il rischio è quello di imitare un altro strumento solo formalmente, poco più che graficamente, senza che cambi la sostanza della ricerca, ovvero di indurre gli utenti dei cataloghi a credere che i meccanismi di ricerca siano omogenei a quelli dei motori di ricerca, che si basano principalmente su metodologie post-booleane, che fanno perno sull'applicazione alla ricerca di metodi probabilistici.

Anche l'approccio e l'uso dei soggetti da parte degli utenti è con tutta probabilità da ripensare, proprio perché si chiede alle biblioteche e ci si aspetta dalle biblioteche, che i loro dati siano da questo punto di vista certificati, rispetto a quelli della rete. Ma forse è anche il concetto di accesso semantico che va espanso: le stringhe di soggetti delle registrazioni catalografiche o le esoteriche numerologie classificatorie non sono più sufficienti a rispondere alle esigenze degli utenti, quando si hanno a disposizione, indici, *abstracts*, *full-text* e altrettanto vale per i metodi di ricerca per soggetto molto limitati di cui disponiamo.

Ancora, e può senz'altro essere una scelta positiva, vengono spesso inseriti elementi del cosiddetto Web sociale negli OPAC, però va tenuto conto che le non moltissime indagini sugli interessi degli utenti degli OPAC, effettuate in anni recenti, non paiono indicare uno specifico interesse verso questi possibili arricchimenti degli OPAC. Così nell'analisi effettuata da Tam, Cox e Bussey di tre indagini sull'uso degli OPAC effettuate in paesi diversi fra studenti di tre università (una statunitense, una di Singapore e una inglese), emerge che le funzionalità ritenute meno interessanti dagli utenti sono l'applicazione di RSS (Really Simple Syndication, formato per la distribuzione di contenuti Web) e la possibilità di in-

<sup>14</sup> Si veda in proposito anche l'accurata analisi in Jia Mi – Cathy Weng, *Revitalizing the library OPAC: interface, searching, and display challenges*, «Information Technology and Libraries», 27 (2008), n. 5, p. 5-22.

<sup>15</sup> «In the post mass digitization era, every word and phrase from millions of digital texts of all literary genres will be at the fingertips of online library catalog users. Giving users a Boolean-based system to search digitized texts is comparable to giving Captain Kirk a Mercury-era space capsule to travel the galaxy» (Markey, *The online library catalog*, cit.).

trodurre commenti o “recensioni”<sup>16</sup>. Non è affatto detto che queste analisi riflettano davvero la realtà e sono per altro limitate a studenti universitari, mentre fra gli utenti di altre tipologie di biblioteche la possibilità di lasciare commenti sui libri letti può essere valutato più interessante quali, ad esempio, quelli per la narrativa nelle biblioteche. Tuttavia è forte il dubbio circa la reale efficacia di esportare modelli di comunicazione da ambiti diversi verso l’OPAC, modelli tra l’altro che si basano in modo predominante su concetti di comunità sociale, che difficilmente si possono adattare al catalogo della biblioteca.

Un esempio, non proprio del nostro settore può essere assai indicativo. Istituzioni come la Library of Congress o la Smithsonian hanno da tempo pubblicato alcuni dei loro moltissimi archivi fotografici su Flickr<sup>17</sup>: un esito interessante è stato quello di ottenere una serie di commenti che, in molti casi, possono aiutare a identificare luoghi e persone non facilmente identificabili, o anche proposte di revisioni e correzioni delle descrizioni standard dei materiali fotografici che rivelano come non sempre ci sia sintonia tra le aspettative degli utenti e l’offerta che istituzioni anche prestigiose riescono a fornire, anche nella catalogazione, che è il più classico dei nostri servizi.

Quindi accanto all’inserimento di modalità di ricerca e presentazione diffusi negli strumenti di ricerca generalisti, può essere altrettanto (o forse più) utile, progettare strumenti che consentano all’OPAC di essere raggiunto da servizi diffusi nella rete, in modo che questa interazione avvenga all’interno dei flussi di lavoro dei nostri utenti: è questa una osservazione di Lorcan Dempsey di qualche anno fa, che resta ancora una guida fondamentale per decidere quali nuove funzionalità valga la pena rendere disponibili nei nostri servizi, non solo negli OPAC<sup>18</sup>. Infatti è esperienza comune l’interesse che si ha nell’utilizzare servizi che bene si integrino nel proprio flusso di lavoro, mentre altri servizi possono assumere il ruolo di “gadget”, anche divertenti e tutt’altro che inutili a priori, ma alla fine poco utilizzati e non percepiti come un effettivo vantaggio. Questa osservazione vale anche per gli utenti delle biblioteche, specie per quelli più vicini e direttamente interessati al catalogo.

Certamente è assai utile incrementare i collegamenti diretti con servizi di rete offerti alla propria comunità, in modo che l’OPAC e i conseguenti servizi della biblio-

<sup>16</sup> Per contro «The most popular features among the three users’ surveys are relevance ranking and borrowing suggestions. Evidently, users want borrowing suggestions for the next-generation OPACs», Winnie Tam – Andrew M. Cox – Andy Bussey, *Student user preferences for features of next*, «Program: Electronic Library and Information Systems», 43 (2009), n. 4, p. 349-374, doi:10.1108/00330330910998020.

<sup>17</sup> Sistema di gestione delle fotografie.

<sup>18</sup> «Historically, library users have adapted their workflow to the library. As the network becomes more important, libraries need to adapt their services to the network workflows of their users», Lorcan Dempsey, *The (digital) library environment: ten years after*, «Ariadne» n. 46, February 2006, <http://www.ariadne.ac.uk/issue46/dempsey/intro.html>.

teca siano integrati con gli altri servizi messi a disposizione degli utenti non troppo remoti, quali l'inserimento nel catalogo degli e-books disponibili in biblioteca o i legami diretti a risorse elettroniche cui la biblioteca è abbonata. È questo un servizio che da tempo abbiamo aggiunto all'OPAC, ma che spesso si fatica a mantenere costantemente aggiornato e funzionale, perché quasi mai i fornitori degli OPAC hanno sviluppato strumenti software davvero efficienti per la gestione di questo tipo di informazione, molto spesso caratterizzata da un alto tasso di variabilità in tempi relativamente ristretti.

Esistono però altri tipi di possibili arricchimenti dell'OPAC in riferimento alle informazioni o utilità presenti in rete.

Così ad esempio, alcuni OPAC statunitensi offrono un *plug-in* che, installato sul browser, consente a un utente della biblioteca che trova un libro sul sito di Amazon di essere automaticamente avvertito nel caso in cui quel libro sia posseduto dalla biblioteca e possa quindi essere preso in prestito.

Oppure, è assai utile in Google Scholar il link ai grandi cataloghi nazionali. Purtroppo manca ancora il link al catalogo nazionale italiano, mentre sono presenti oltre a WorldCat, vari cataloghi nazionali, quali lo svedese Libris, lo spagnolo Rebiun e altri. È attivo già da tempo, ed è senz'altro assai utile, il link all'ACNP (Archivio collettivo nazionale dei periodici) ma poiché non poche sono le monografie che oramai sono presenti in Google Scholar, sarebbe senz'altro utile per gli utenti avere il link diretto a SBN (Servizio bibliotecario nazionale).

O ancora, sono pochi gli OPAC italiani che consentono di salvare i dati bibliografici in Zotero, il *plug-in* per client Web specializzato per la creazione e la gestione di bibliografie personali e che è oramai assai diffuso presso le comunità scientifiche.

A proposito di queste tematiche mi sembra assolutamente condivisibile l'osservazione di Metitieri, in chiusura di uno dei suoi ultimi contributi pubblicati:

«l'OPAC ha bisogno di essere rivisto, quindi una sua semplificazione sarebbe molto utile e anche la sperimentazione di meccanismi di collaborazione con gli utenti risulterebbe interessante. Con un progetto, però, e verificandone i risultati passo per passo. senza vuote ideologie e soprattutto mettendo da parte gli entusiasmi ingiustificati»<sup>19</sup>.

Emergono insomma due possibili tendenze di evoluzione degli OPAC che possono apparire diametralmente opposte. Da un lato un catalogo, ma più generalmente un servizio di biblioteca, può essere utilizzato da utenti remoti se è in grado di raggiungere una reale massa critica, obiettivo che è realizzabile solo nell'ambito di cooperazioni allargate e di portata sempre più sovranazionali, anche sacrificando parte della propria identità locale. E comunque gli investimenti complessivi che l'innovazione tecnologica richiede, in termini non solo di costi diretti, ma anche di

<sup>19</sup> Fabio Metitieri, *L'OPAC collaborativo, tra folksonomia e socialità*, «Biblioteche oggi», 27 (2009), n. 2, p. 12.

competenze e risorse umane, impongono sempre più di agire all'interno di cooperazioni allargate, se si desidera essere punti di riferimento nella rete.

D'altro canto per rispondere al meglio alle esigenze degli utenti locali è necessario rivedere più di un aspetto degli OPAC attuali, probabilmente incidendo anche su alcuni aspetti di base, senza contare la problematica enorme, che qui neppure si è citata, che pone la ormai evidente obsolescenza dei formati in cui i dati sono raccolti e gestiti nei cataloghi. Tali revisioni, per tentare di rispondere nel modo più adeguato alle esigenze della utenza locale, dovrebbero puntare a una forte identificazione e riconoscibilità della biblioteca e della sua collezione specifica, aspetti che sono spesso trascurati dagli OPAC attuali.

Sono suggestioni opposte, che rispondono a obiettivi diversi, ma che comunque comportano investimenti importanti in termini non solo di costi, ma anche di fantasia, di capacità di innovare, di condividere con comunità sempre più larghe obiettivi, esperienze, e – perché, no – sogni.

*OPACs are evolving towards a new generation of query interfaces, under the pressure of both automated system producers and libraries. Despite there now being more and richer digital collections available on the net, the way such query interfaces should evolve is not yet clear, particularly with reference to the users these interfaces are supposed to serve. In order to reach out for remote online users, participation to international and world catalogues has become a must. Competing with generic query engines will indeed only be possible if a critical mass of data of significant size will be created, with the support of remarkable investments. Nevertheless, while rethinking OPACs we should not forget the needs of on site library users, to whom the catalogue provides access to the specific physical or digital collection represented in the catalogue itself. For libraries, the evolution of OPACs thus represents an important challenge – an opportunity that must be grasped but whose objectives should be clearly defined.*

# Il paradosso della conservazione digitale: riflessioni sull'autenticità\*

**Giovanni Michetti**

Università degli studi di Roma "La Sapienza"

*L'autenticità è uno dei nodi cruciali della conservazione a lungo termine degli oggetti digitali: nonostante gli sforzi e le iniziative di ricerca promosse a livello internazionale, la comunità scientifica non dispone oggi di una teoria compiuta dell'autenticità in ambiente digitale, pur potendo fare riferimento ad un corpus di analisi, riflessioni e materiali ormai consolidato. Il presente articolo si inserisce nel dibattito in materia e, prendendo le mosse dal carattere paradossale della conservazione digitale, in bilico fra le opposte esigenze di trasformazione e di non alterazione degli oggetti, suggerisce degli elementi di riflessione per una migliore definizione del concetto di autenticità. Sulla base delle considerazioni introdotte, viene poi illustrato sommariamente il modello di autenticità elaborato all'interno del progetto europeo CASPAR e sviluppato coerentemente con l'approccio concettuale proposto nella prima parte del saggio.*

L'autenticità è sicuramente un nodo cruciale della conservazione a lungo termine degli oggetti digitali: il ruolo dell'autenticità come requisito preliminare e fondamentale dei processi conservativi è stato analizzato all'interno di vari progetti internazionali, alcuni concentrati sulla conservazione a lungo termine di oggetti digitali nell'ambito scientifico, culturale e della pubblica amministrazione (primo fra tutti InterPARES<sup>1</sup>), altri dedicati all'identificazione dei criteri e delle responsabilità per lo sviluppo di depositi digitali affidabili (come ad esempio il progetto tedesco NESTOR<sup>2</sup> o l'iniziativa congiunta del Research Library Group e del National Archives and Records Administration<sup>3</sup> negli Stati Uniti).

Uno dei vincoli fondamentali per lo sviluppo di una teoria dell'*autenticità digitale* risiede nell'impossibilità pratica di conservare le risorse digitali nelle loro condizioni originali<sup>4</sup>, senza alcuna alterazione: nella maggior parte dei casi siamo costretti ad un'incessante attività di migrazione degli oggetti che non può essere ridotta ad un semplice spostamento su nuovi supporti, ma che si configura piuttosto come

\* Il testo è una rielaborazione dell'intervento presentato al convegno "I luoghi delle memorie e della conoscenza" (Ravenna, Archivio di Stato, 14 novembre 2008).

<sup>1</sup> <http://www.interpares.org>.

<sup>2</sup> <http://www.langzeitarchivierung.de/eng/index.htm>.

<sup>3</sup> <http://www.oclc.org/research/activities/past/rlg/repositorycert.htm>.

<sup>4</sup> Si è preferito fare riferimento alle *condizioni* piuttosto che alla forma o allo stato, onde esprimere nella maniera più generale possibile l'insieme delle caratteristiche fisiche e logiche o – se si preferisce – dei caratteri estrinseci ed intrinseci di un oggetto.

una vera e propria trasformazione, in maniera tale che possiamo solo *ri-produrre* gli oggetti originali o – meglio ancora – possiamo solo riprodurre la percezione degli oggetti originali. In breve: non possiamo conservare gli *originali*<sup>5</sup>. Sfortunatamente questa circostanza abbastanza ovvia muove in senso contrario ad una fondamentale assunzione di principio, e cioè che la conservazione dell'autenticità implica la conservazione dell'*identità* e dell'*integrità* dell'oggetto digitale. Non è neppure necessario scomodare la dimensione tecnologica del problema o ricorrere a sofisticate considerazioni teoretiche: il dato di esperienza ci suggerisce di ritenere autentico un oggetto che non sia stato modificato o corrotto, *in primis* nelle sue caratteristiche fisiche (integrità), ma non secondariamente anche in relazione alla sua natura, al contesto di appartenenza o al suo profilo logico (identità). Si genera così una sorta di paradosso – ma neppure tanto, a ben pensarci – ove gli oggetti sono costretti a cambiare per rimanere se stessi: l'obsolescenza tecnologica impone il cambiamento, lo spostamento, il fattore dinamico insomma; l'autenticità invece richiede – o meglio, sembra richiedere – la staticità, la *fissità* degli oggetti. In altre parole, una sorta di *cristallizzazione* della fonte sembrerebbe l'espediente più sicuro per garantire ai posteri la genuinità della fonte stessa, intesa non solo come protezione nei confronti delle eventuali corrotte determinate dalla temperie del tempo, ma anche come consolidamento e stabilizzazione delle caratteristiche complessive dell'oggetto stesso. La conservazione riguarda infatti l'insieme delle caratteristiche di un oggetto, e non *sic et simpliciter* l'oggetto stesso: è un lieve slittamento nella formula linguistica che però sottintende un profondo convincimento culturale e metodologico. Nonostante le apparenze, la conservazione – intesa nella sua accezione più completa e complessa – non è mai rivolta alla sola materialità di un oggetto, sia esso documentale o di altro tipo: la conservazione riguarda l'oggetto *tout court*. Ciò ha significato – fino ad oggi – preservarne la materialità, poiché questa si è posta come ineludibile strumento e veicolo del contenuto informativo dell'oggetto stesso. In altri termini, anche in ambiente tradizionale la conservazione della materialità non è il fine, ma piuttosto il mezzo attraverso il quale noi conserviamo il patrimonio di valori (siano essi informativi, ideali, tecnici o latamente culturali) soggiacenti all'oggetto stesso. Questa prospettiva non esclude ovviamente che vi siano ambiti nei quali la materialità *in sé* comunichi valori degni di conservazione: basti pensare alle opere d'arte o – per restare nel nostro dominio – alle fonti documentarie di epoche lontane, ove la fattura stessa del documento è un valore, se non il valore preminente, e il documento porta inciso sulla sua pelle, metaforicamente e concretamente, un significato che

<sup>5</sup> Ovviamente, come evidenziato dall'uso del corsivo, si fa qui riferimento alla nozione tradizionale di *originale*, giacché l'ambiente digitale impone a nostro avviso un ripensamento di questa categoria concettuale e l'adozione di un mutato paradigma che consenta di considerare originali anche gli oggetti che abbiano subito delle trasformazioni, a patto che queste non modifichino le proprietà significative degli oggetti stessi.

possiamo per lo meno qualificare come tecnico, ma che qualunque studioso della cultura scritta non esisterà a riconoscere come culturale. Ma non si può non riconoscere che l'ambiente digitale sembra proporci una prospettiva nuova: la conservazione del bene culturale non impone la conservazione della materia originale<sup>6</sup>.

A tutto ciò si aggiunga che, a ben guardare, anche in ambiente tradizionale la conservazione non si è mai limitata alla sola materialità, ma si è estesa all'insieme delle conoscenze che fungono da necessario corredo per la corretta interpretazione di una risorsa: senza una conservazione delle idealità soggiacenti e del contesto culturale, antropologico, tecnico, etc. entro cui l'oggetto è immerso, non riusciremmo oggi a recuperare il *patrimonio culturale*, cioè un insieme di idealità trasmutate nella materialità, e avremmo un vuoto oggetto incapace di comunicarci i suoi significati. In altre parole, non occorre dimenticare che – in maniera implicita o esplicita – insieme agli oggetti abbiamo sempre conservato un *corpus* di conoscenze che in una certa misura è parte dell'oggetto stesso, poiché partecipa del suo sistema di significati.

Il paradosso fra staticità e dinamismo si ripercuote sulla *stabilità* dell'autenticità in ambiente digitale: questa non può essere riconosciuta come data una volta per tutte, poiché è inficiata ogni qual volta un oggetto digitale è trasferito nello spazio (ad esempio nello scambio fra utenti, sistemi, applicativi) o nel tempo (per semplice permanenza sul supporto di archiviazione o per migrazione, aggiornamento o in generale trasferimento).

L'estrema labilità e l'evidente incertezza dei contorni dell'autenticità impongono

<sup>6</sup> Beninteso, la storia delle successive migrazioni da un supporto all'altro deve essere conservata, perché ogni supporto – e ogni sistema logico entro il quale l'oggetto è immerso – ha delle specifiche caratteristiche. Il rigore filologico impone di tenere traccia di questo "percorso", non foss'altro perché il supporto stesso ci comunica informazione: una gestione con cd-rom e juke-box è più adeguata per una conservazione cosiddetta *near-line* o *off-line*; l'uso odierno di un floppy disk denota probabilmente un ritardo tecnologico; una chiave USB è riservata in linea di massima a duplicazioni o spostamenti temporanei di dati, non certo alla conservazione permanente; e così via. Senza calcolare che il passaggio da un supporto all'altro, o fra due supporti dello stesso tipo diversamente formattati, potrebbe causare perdita d'informazione. Insomma, l'elemento materiale non è trascurabile, per lo meno non sempre. Il punto è che in ambiente digitale il suo rilievo è in genere assolutamente minimo, e men che mai paragonabile al suo *status* nell'ambiente tradizionale. Molto concretamente: non è affatto improbabile che un *file*, memorizzato in origine su un floppy disk, giunga fino a noi sotto forma di *file* archiviato su disco fisso. Tale circostanza non sembra problematica: la conoscenza del supporto originale serve certo per immaginare le modalità di fruizione della risorsa (e cioè reperimento del disco, inserimento nel drive, tempi di accesso più lenti, maggiore sensibilità al degrado, ecc.); e la conoscenza di un'avvenuta trasformazione deve renderci consapevoli della possibilità di perdita di dati rilevanti. Ma il cuore della fruizione rimarrebbe inalterato: ci ritroveremmo (sempre che sia stato conservato tutto l'apparato tecnologico necessario e che si possa accedere all'oggetto senza difficoltà) davanti a uno schermo e alla rappresentazione dell'oggetto. Per correttezza filologica potremmo disquisire sul fatto che tale rappresentazione è oggi realizzata grazie a schermi ultratecnologici e iperdimensionati, mentre l'utente del passato disponeva ahilui di un ingombrante terminale VT100 con un piccolo video in bianco e nero; ma ci addentreremmo in un livello di analisi più sofisticato, ove è bene che esercitino le proprie competenze altre discipline.

l'adozione di misure a sostegno di una gestione controllata della risorsa: l'autenticità di una risorsa deve essere comprovata e sostenuta da *prove* associate alla risorsa stessa, attraverso una qualsivoglia forma di *documentazione* che tracci e testimoni la storia dei processi di migrazione e di elaborazione, in una parola dei processi di *trasformazione* dell'oggetto nel corso del tempo. In particolare, occorre documentare l'utilizzo di strumenti, tecniche e strategie mirate a garantire l'*identità* e l'*integrità* delle risorse, o almeno in grado di minimizzare i rischi di cambiamento nel corso del tempo.

La conseguenza di tale approccio è che l'autenticità non è mai limitata alla risorsa oggetto di analisi, non è una qualità della risorsa in sé, ma si estende all'intero sistema informativo/documentario e alla sua affidabilità. Ed è per questo motivo che l'autenticità implica un controllo continuativo tanto del contesto di produzione quanto dell'ambiente di trattamento e conservazione della risorsa.

Non possiamo tuttavia nascondere che tali considerazioni, pur ragionevoli e fondate sulle acquisizioni teoriche di importanti progetti di ricerca internazionali, presentano un *vulnus* che abbiamo finora sottratto volutamente all'attenzione del lettore: in assenza di una chiara e inequivocabile definizione del concetto di autenticità, ogni ragionamento sul tema sarà caratterizzato da un'intrinseca debolezza che inficia la costruzione di una teoria in grado di autosostenersi. Ed è per questo che nelle battute iniziali del nostro saggio abbiamo fatto riferimento al dato di esperienza<sup>7</sup>.

Cerchiamo quindi di approfondire il concetto di autenticità indagandone le fondamenta: secondo la definizione più autorevole e consolidata,

«un documento è autentico se è ciò che dichiara/mostra di essere ed è esente da alterazioni o corruzioni»<sup>8</sup>.

Tale definizione, benché sufficientemente generica da tollerare diverse interpretazioni, ad una lettura più approfondita risulta eccessivamente vaga, e appare insoddisfacente anche in ambito digitale. L'analisi puntuale della formula definitoria solleva delle criticità che vale la pena assumere come materia di ragionamento intorno a questi temi, con l'auspicio di raffinare ulteriormente il concetto di autenticità.

<sup>7</sup> A onor del vero, anche il profilo esperienziale è tutt'altro che pacifico se solo si consideri che l'autenticità, ben lungi dall'essere una caratteristica immanente o latente degli oggetti, può essere interpretata come fenomeno sociale e, come tale, storicizzato. Tuttavia in questa sede non interessa tanto illustrare le diverse prospettive di analisi del concetto di autenticità, quanto rilevare alcune caratteristiche che ne evidenzino la complessa natura.

<sup>8</sup> «Authenticity [is] the quality of a record that is what it purports to be and that is free from tampering or corruption». Cfr. InterPARES 2 Project, *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*, a cura di Luciana Duranti e Randy Preston, Roma: ANAI, 2008, p. 775, disponibile all'indirizzo [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_book\\_complete.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_book_complete.pdf).

1. Posta in questi termini, l'introduzione fraudolenta di un documento autentico – di più: giuridicamente autentico – all'interno di un sistema informativo produce un documento autentico? Apparentemente sì, purché il documento sia ciò che dichiara di essere e sia esente da manipolazioni. Ma rispetto alla nostra interpretazione di autenticità<sup>9</sup>, rispetto all'esigenza di riconoscere l'identità di un oggetto ben oltre i confini dell'oggetto stesso, rispetto al ruolo dell'*ambiente* che dà forma e sostanza (cioè, significato) all'oggetto: rispetto a tutto ciò, il documento introdotto fraudolentemente non è un documento autentico, perché basta estendere lo sguardo al di là del semplice oggetto e considerare il *contesto* di produzione per verificare presumibilmente la presenza di una falla nel sistema documentario, nell'insieme dei processi che hanno consentito a quell'oggetto di essere assunto nel sistema senza averne diritto, non rispettando cioè quelle procedure che complessivamente garantiscono l'affidabilità del sistema a tutela dell'autenticità degli oggetti. Il punto è quindi interpretare in maniera assolutamente estensiva la locuzione "essere ciò che dichiara di essere" o – meglio ancora – riformularla, utilizzando ad esempio una formula del tipo: "un documento è autentico se è ciò che *manifesta* di essere", ove tale *presentazione verso l'esterno* sia riferita non solo al contenuto informativo del documento, ma anche al complesso intreccio di relazioni che legano il documento ad un contesto. In questo secondo caso, il documento introdotto fraudolentemente nel sistema non passerebbe il vaglio della definizione, poiché si propone evidentemente non solo come documento caratterizzato da una sua autonomia e da una sua intima natura (documento di debito, contratto, circolare; sottoscritto da Tizio o da Caio in una certa data; ecc.), ma come documento appartenente ad un sistema di relazioni, si *manifesta* cioè come documento del sistema informativo, e ciò sarebbe in contrasto con la realtà dei fatti. Sarebbe – ci si consenta – un falso non tanto dal punto di vista giuridico, quanto dal punto di vista archivistico.
2. Nella definizione, il generico riferimento all'assenza di alterazioni e corruzioni rinvia a quell'ipotesi di cristallizzazione degli oggetti che però abbiamo già escluso dal nostro orizzonte per evidenti motivi tecnici. Di nuovo, evitando di interpretare rigidamente la definizione onde salvarne lo spirito, la formula potrebbe essere riferita agli aspetti rilevanti – o meglio, essenziali – per garantire l'autenticità. Ma così si rischia evidentemente di cadere in una tautologia: un documento è autentico se è esente da alterazioni o corruzioni riferite agli aspetti rilevanti per l'autenticità. La nostra impressione è che il generico riferimento ad alterazioni o corruzioni sia inadeguato e convenga abbandonarlo piuttosto che cercare di modificarlo. Occorre riconoscere che il coraggioso

<sup>9</sup> *Ibidem*.

tentativo del progetto InterPARES di trasportare in ambiente digitale concetti e metodi delle discipline tradizionali ha avuto valore più per il consistente *corpus* di analisi, riflessioni, materiali prodotti sull'argomento, che non per l'elaborazione di una teoria compiuta dell'autenticità in ambiente digitale. La profondità dei livelli di mediazione imposti dagli oggetti digitali e l'enorme varietà dei formati e delle strutture di dati sono tali che non risulta praticabile l'individuazione a priori di categorie concettuali, di profili, di caratteristiche meritevoli di attenzione specifica. Ad esempio, il degrado di un'immagine attraverso un'operazione di compressione che ne riduca la qualità grafica può essere inaccettabile in alcuni contesti, mentre in generale ciò non vale per un oggetto testuale, la cui qualità grafica è vincolata a soglie di tolleranza ben più ampie (non ci riferiamo ovviamente alla digitalizzazione di una pergamena). Una stessa operazione, quindi, ha ripercussioni diverse su oggetti diversi, perché in alcuni casi l'oggetto potrebbe non essere più riconosciuto come autentico. In altre parole: la qualità grafica è una caratteristica che ha un valore costitutivo e funzionale diverso a seconda dei casi. Ed è pertanto impossibile delinearne a priori il valore in riferimento all'autenticità. Ergo, la costruzione di una teoria dell'autenticità sarà sempre costretta ad assumere connotati di *genericità* e *rarefazione* tali da renderla – ci si passi la provocazione – inutilizzabile se non come quadro generale, come impianto teorico da cui derivare specifiche teorie, metodologie e applicazioni riferibili a specifici contesti e oggetti.

3. A ben pensarci – ulteriore apparente paradosso – ogni oggetto è in sé autentico e integro. Ciò che sembra mancare nella definizione di InterPARES è il legame con la risorsa primigenia: in altri termini, il concetto di autenticità è intimamente legato, anche se non appare all'evidenza, ad un profilo diacronico, ad un movimento nel tempo, ad un'idea quasi deterministica di individuazione del punto iniziale A da cui prendono le mosse i nostri ragionamenti (i fisici direbbero "dal punto  $t_0$ ") e di un punto finale B ( $t_1$ ). E volutamente usiamo il termine *punto*, intendendo con ciò non solo un istante nel tempo, ma un preciso ente dello spazio-tempo, cioè un evento che deve essere contestualizzato nel suo ambiente di produzione. È vero: le "alterazioni e corruzioni" di cui alla definizione assunta a modello evidentemente rimandano a questo profilo dinamico, ma oggettivamente si tratta di un richiamo piuttosto debole. Le conseguenze di tale approccio *dinamico* possono essere tutt'altro che scontate: in una visione per così dire *statica* noi concentriamo la nostra attenzione su un oggetto e ne seguiamo l'evoluzione verificando – o meglio, cercando di verificare – ad ogni generazione successiva, ad ogni passaggio cruciale nel ciclo di vita del documento, che i caratteri essenziali del documento permangano in esso, così da poter riconoscere nel nuovo oggetto l'antico. Accettiamo cioè una catena ininterrotta di trasformazioni che ci conducono dall'oggetto

$A_1$  all'oggetto  $A_n$  con tutta probabilità perdendo progressivamente traccia dell'oggetto iniziale, ma confidando nelle procedure che ci hanno condotto fino all'oggetto  $A_n$  e che garantiscono la sostanziale equivalenza con l'oggetto  $A_1$  dal punto di vista dell'autenticità<sup>10</sup>.

In una visione più *dinamica* e fortemente fondata sul concetto di tempo e stati (in relazione al tempo), l'oggetto iniziale è solo l'inesco per una catena di trasformazioni che danno vita a nuovi oggetti, ognuno con una propria dignità: in questo caso, la migliore astrazione per rappresentare queste vicende è un grafo, una sorta di *stemma codicum* che racconti la storia dell'oggetto e che in maniera più neutrale descriva le trasformazioni non fornendo necessariamente garanzie sull'autenticità. La conseguenza immediata di questo approccio è ovviamente la necessità di conservare tutti gli oggetti della catena.

In linea di massima, tanto nell'ambito sociale quanto nei domini di settore finora si è preferito adottare la prima interpretazione, tutta concentrata sull'oggetto e sulla permanenza dei suoi caratteri significativi. In quest'ottica, si potrebbe dunque assumere l'integrità e l'identità di una risorsa come discriminare per l'attribuzione dell'autenticità, il che consentirebbe di proporre una definizione di autenticità più limpida, nella quale sia assorbito anche il profilo dinamico: una risorsa rimane autentica nel corso delle successive trasformazioni se conserva l'integrità e l'identità. Evidentemente, tale definizione non è in sé conclusa poiché rinvia ad altri concetti, ma presenta il duplice vantaggio di assorbire esplicitamente il profilo dinamico (le successive trasformazioni), e di spostare il *focus* del problema sulle due qualità cruciali dell'oggetto (integrità e identità), che pertanto richiedono – se non un'esplicita definizione – un adeguato approfondimento<sup>11</sup>.

L'*integrità* di una risorsa si riferisce alla sua completezza (*wholeness*): una risorsa è integra quando è completa ed esente da corruzioni nei suoi aspetti fondamentali. Il processo di verifica dell'autenticità dovrebbe analizzare proprio tali aspetti e accertare che resistano agli inevitabili cambiamenti indotti dall'obsolescenza tecnologica. Per essere più concreti, nei processi conservativi il mantenimento del flusso di bit non è quasi mai necessario, mentre è inderogabile la conservazione e la completezza della cosiddetta *forma intellettuale*, con particolare riferimento a quelle caratteristiche che veicolano significati di rilievo. In altre parole, l'integrità

<sup>10</sup> Di fronte ad una copia autenticata da un notaio non riserviamo interesse alcuno alle modalità con cui avviene il processo di autenticazione, confidando nel ruolo di terzietà del pubblico ufficiale e nella sua capacità di produrre documentazione fidefacente. Non importano i meccanismi: ciò che rileva è il risultato finale, un prodotto che – sotto il profilo giuridico – possa ritenersi equivalente all'originale in forza di una valutazione demandata ad un soggetto in grado di cogliere gli aspetti rilevanti del documento e verificarne la permanenza, con tutte le garanzie proprie della sua *auctoritas*.

<sup>11</sup> Le nozioni di integrità e identità presentate di seguito affondano le loro radici nel quadro concettuale elaborato all'interno del progetto InterPARES.

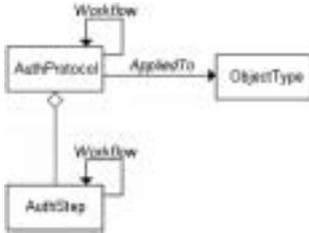
fisica di una risorsa (il *bitstream* originale) può essere corrotta, ma la struttura del contenuto e le componenti essenziali devono rimanere intatte.

L'integrità cioè non è mai totale: c'è sempre un grado di deterioramento. Quindi il problema cruciale è identificare le caratteristiche rilevanti, quelle che definiscono la *forma intellettuale*: questo significa comprendere la natura della risorsa, analizzarne le caratteristiche e valutarne la loro funzione in maniera tale da stabilire quali cambiamenti sono ammessi, senza timore che inficino l'integrità.

L'*identità* di una risorsa deve essere intesa in un'accezione molto ampia: essa si riferisce non solo alla sua univoca denominazione e identificazione. L'identità si riferisce all'insieme delle caratteristiche di una risorsa che univocamente la identificano e la distinguono da tutte le altre: si riferisce cioè non solo alla sua struttura concettuale interna, ma anche al suo contesto generale, cioè al sistema di relazioni cui partecipa (siano esse amministrative, legali, documentali, tecnologiche, perfino sociali). Una risorsa non è una monade isolata, con dei confini ben definiti e una vita propria: una risorsa è un oggetto *nel contesto*, è l'oggetto stesso e l'insieme delle relazioni che danno significato all'oggetto (riferendoci alle considerazioni precedentemente esposte, non è un punto isolato, bensì un punto dello spazio-tempo). Se vogliamo, si tratta di una banale considerazione epistemologica: conoscere gli oggetti significa conoscere un *intorno* di quegli oggetti, sufficientemente ampio da ricavarne un livello di significatività ritenuto discrezionalmente sufficiente. Il problema è che queste relazioni cambiano nel corso del tempo e quindi abbiamo bisogno non solo di comprenderle ed esplicitarle, ma anche di documentarle al fine di avere una *storia* completa della risorsa: non possiamo perdere questa *storia* senza perdere un po' dell'identità della risorsa, con conseguenze (negative) sulla valutazione dell'autenticità della risorsa stessa.

A proposito di valutazione: occorre fare una chiara distinzione fra l'autenticità di una risorsa e la procedura di *validazione* della risorsa, di verifica della sua autenticità. La seconda è parte di un processo più generale mirato ad assicurare che l'oggetto informativo sia conservato *come se, in luogo di*, insomma *funga da* originale. Gli strumenti di gestione dell'autenticità devono quindi monitorare e gestire protocolli e procedure all'interno della catena della custodia, dalla fase di creazione lungo tutto il processo di conservazione. L'autenticità non può essere valutata per mezzo di un indicatore booleano che ci dica se la risorsa è autentica o meno. Nonostante sembri strano, nella valutazione esistono dei *gradi di autenticità*: la certezza dell'autenticità di una risorsa è un obiettivo e i casi certi sono casi limite. La valutazione *tende asintoticamente all'autenticità*, in funzione del grado di controllo di oggetti e processi, e quindi dobbiamo predisporre meccanismi e strumenti conservativi ricordando sempre che alterazioni, corruzioni, perdite di dati significativi e altre modifiche interne o esterne sono dietro l'angolo, anzi sono una certezza; dobbiamo cioè progettare strumenti e *pesi* per comprendere il rilievo di tali cambiamenti e il loro impatto sull'autenticità. Continuando a svolgere il filo del paradosso, dobbiamo fare della probabilità e della statistica il terreno delle certezze sui cui rifondare taluni metodi.

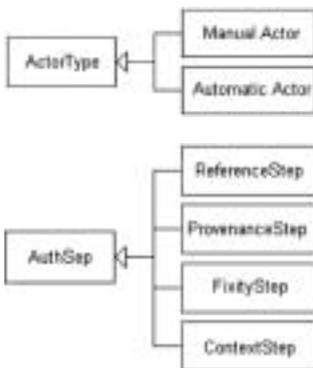
La conseguenza è che gestire l'autenticità implica operare su un insieme appropriato di attributi relativi a contenuto e contesto, e verificare (magari con l'ausilio di un'apposita metrica) la completezza o il grado di alterazione di questo insieme. Nel progetto europeo CASPAR è stato tentato un approccio basato su queste riflessioni<sup>12</sup>. L'osservazione cruciale da cui siamo partiti è che la protezione dell'autenticità e la sua valutazione costituiscono un *processo*. Per gestire tale processo abbiamo bisogno di definire le procedure che occorre seguire in relazione a specifiche tipologie di oggetti e di eventi. Abbiamo definito una tale procedura Authenticity Protocol (AP).



Un AP è un insieme di passi fra loro collegati, ognuno dei quali prende il nome di Authenticity Step (AS). Ogni *step* modella una parte del protocollo (AP) e può essere eseguito autonomamente, come un processo autonomo; l'Authenticity Step (AS) costituisce in pratica una fase del processo globale mirato alla valutazione dell'oggetto. Le relazioni fra i vari passi stabiliscono in che ordine questi debbano essere eseguiti nel contesto di uno specifico protocollo. Senza entrare nei dettagli, denotiamo semplicemente con Workflow l'insieme di tali relazioni. A sua volta, ogni Authenticity Protocol può essere usato ricorsivamente per definire altri protocolli, come rappresentato dalla presenza della relazione di Workflow.

Il protocollo di autenticità si applica a un ObjectType, una classe di oggetti con caratteristiche uniformi per l'applicazione del protocollo.

Uno *step* è eseguito da un ActorType, sia esso automatico (hardware o software) o manuale (ente o persona), istanziato da un Actor.

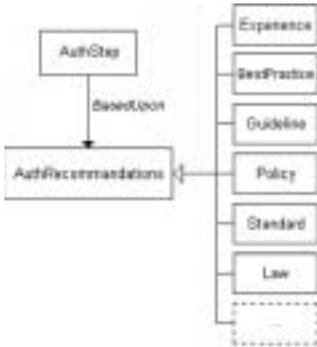


Ci possono essere vari tipi di Authenticity Step: coerentemente con le indicazioni del modello OAIS<sup>13</sup>, distinguiamo gli *step* sulla base delle diverse tipologie di Informazioni sulla conservazione, e quindi avremo *step* riferibili all'identificazione (*reference step*), *step* riferibili alla provenienza (*provenance step*), altri riferibili all'integrità (*fixity step*) e altri ancora riferibili al contesto (*context step*).

Qualunque analisi svolta sull'oggetto può essere riferita a uno di questi *step* o ad una loro combinazione.

<sup>12</sup> Il sito ufficiale del progetto è all'indirizzo <http://www.casparpreserves.eu>. L'Autore ha partecipato al progetto di ricerca come membro del team dell'Università di Urbino che ha collegialmente sviluppato tale modello, con il contributo scientifico dell'ISTI (CNR Pisa) e di Engineering Ingegneria Informatica.

<sup>13</sup> OAIS: *sistema informativo aperto per l'archiviazione*, a cura di Giovanni Michetti, Roma: ICCU, 2007.



Uno *step* implica un'analisi per una successiva valutazione e pertanto sono necessarie informazioni relative a:

- *best practice*, metodologie e ogni tipo di regolamentazione che deve essere seguita o che possa aiutare nell'analisi e nella valutazione;
- criteri che devono essere eventualmente soddisfatti nella valutazione.

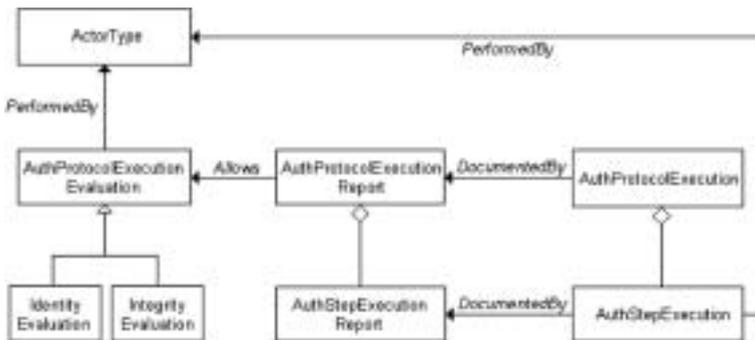


Gli Authenticity Protocol sono eseguiti da un attore su oggetti che appartengono ad una specifica tipologia nel contesto di un'Authenticity Protocol Execution Session. L'esecuzione di un AP è modellata come un Authenticity Execution Protocol (APE). Un APE è correlata ad un AP mediante la relazione ExecutionOf e consiste di vari Authenticity Execution Steps (ASE). Ogni ASE, a sua volta, è correlata ad un AS mediante un'associazione di tipo ExecutionOf, e contiene le informazioni relative all'esecuzione, compresi:

- l'attore che ha fatto l'esecuzione;
- l'informazione utilizzata;
- luogo, tempo e contesto di esecuzione.

Tipi differenti di Authenticity Step Execution hanno strutture differenti e i risultati delle esecuzioni devono essere documentati al fine di raccogliere informazioni su specifici aspetti della risorsa (ad esempio denominazione, consistenza, date e trasformazioni).

Un Authenticity Step Execution Report documenta semplicemente che lo *step* è stato eseguito e raccoglie tutti i valori associati ai metadati analizzati nel corso





Il tema dell'autenticità non può essere certo esaurito in queste poche pagine. Non è stato affrontato ad esempio il profilo del *comportamento*: conservare gli oggetti digitali significa conservare degli oggetti che hanno appunto un comportamento, delle dinamiche. L'autenticità gioca un ruolo cruciale anche in questo ambito, giacché un appiattimento dei fattori dinamici di un oggetto è in contrasto – in generale – con la permanenza dei caratteri di identità dell'oggetto.

Non abbiamo affrontato gli aspetti *organizzativi*: molti ritengono la firma digitale un utile strumento di supporto alle strategie di conservazione e verifica dell'autenticità delle risorse digitali, ma in che misura tale tecnologia può essere d'aiuto? e – soprattutto – qual è il peso della costruzione di una Public Key Infrastructure (PKI) che fornisca affidabilità a tale tecnologia? è possibile prevedere dei modelli organizzativi di rango locale – ad esempio delle infrastrutture di certificazione a livello di ente o di *network* – preservando l'affidabilità di questo strumento<sup>14</sup>? e come gestire concettualmente e praticamente una gerarchia dei certificatori? è ipotizzabile una *gerarchia* dell'autenticità? nel caso, come dovremmo differenziare la gestione e la conservazione delle fonti in funzione della loro autenticità?

Non è stato affrontato neppure il profilo *politico*, sovrapposto in parte a quello organizzativo: come ormai consolidato nella letteratura di settore, i depositi digitali certificati necessitano del ruolo del conservatore come di una terza parte fidata, una terza parte cioè che:

- non abbia un interesse soggettivo sui documenti<sup>15</sup>;
- non abbia ragioni specifiche per alterare la documentazione oggetto di custodia;
- non consenta ad alcuno di alterare la documentazione, accidentalmente o di proposito.

Come garantire tutto ciò non solo dal punto di vista tecnico, ma soprattutto dal punto di vista *politico*? Come individuare l'architettura complessiva e le figure specifiche che nel nostro ordinamento potrebbero assumere questo ruolo? È vero: esiste una normativa di settore; ed esiste una fitta rete di istituti di conservazione sul nostro territorio. Ma il mutato contesto sociale, le mutate esigenze tecniche, perfino le nuove configurazioni istituzionali e costituzionali richiedono un ripensamento di queste logiche per non fotocopiare semplicemente in ambiente digitale strutture e metodi dell'ambiente tradizionale. Sono cioè necessari interventi di carattere politico-organizzativo, come ad esempio:

<sup>14</sup> La nostra osservazione ha un carattere generale: il legislatore italiano ha già dato una risposta a tale quesito, elaborando un complesso *corpus* di norme – in verità niente affatto scevro da importanti criticità interpretative – sulla firma digitale e sui relativi processi di certificazione.

<sup>15</sup> Efficacemente, in letteratura si parla talvolta di "soggetti che non abbiano una *quota* di partecipazione (*stake*) nei documenti".

- una chiara definizione dei ruoli e delle responsabilità in sede conservativa;
- uno sviluppo coerente di raccomandazioni e politiche per la costruzione di depositi affidabili;
- una precisa identificazione delle singole componenti della funzione conservativa.

Tutto ciò, coerentemente con una chiara definizione degli elementi concettuali sui quali fondare la complessa architettura che sostenga la *mission* conservativa in ambiente digitale.

In breve, i profili d'indagine sono molteplici: in questa sede noi abbiamo inteso solo proporre degli spunti di riflessione, evidenziando la necessità di una rigorosa analisi e di un solido impianto concettuale e metodologico per elaborare un modello grafico-simbolico che sorregga le nostre astrazioni e passi il vaglio dell'implementazione. Viceversa, la rappresentazione grafica del modello, basata sulla notazione formale dei linguaggi di modellazione, non solo conferma l'utilità di metodi e tecniche appartenenti a settori disciplinari diversi dal nostro, ma soprattutto rivela che tali costruzioni simboliche sono veicolo di un'idea, di un preciso approccio e di una determinata interpretazione del mondo, soggiacenti a schemi niente affatto neutri.

L'autenticità richiede una riflessione articolata, uno sforzo in più direzioni, una visione ad ampio raggio<sup>16</sup>: non si tratta più di *tradurre* soluzioni consolidate dall'ambiente tradizionale a quello digitale, senza soluzione di continuità. Qui si tratta di *reinterpretare*, rinnovare, cambiare punto d'osservazione. Se occorre, con un pizzico d'incoscienza.

*Authenticity is one of the critical aspects in long-term digital preservation. Despite all efforts and research initiatives that have been promoted at the international level, the scientific community cannot yet rely on the existence of a thorough theory on authenticity in a digital environment. On the other hand, an overall framework of reference, made of analytical tools, critical understanding and reference materials, does exist and is by now fairly consolidated. The present paper explores the debate on this issue and, starting from the paradox of digital preservation – constantly thorn by the conflicting need to transform the objects without modifying them – highlights a number of critical aspects that should be considered in order to better define authenticity as a notion. The paper then follows up to its suggestions by briefly illustrating the authenticity model developed by the EU CASPAR project – a model which well adheres to the theoretical approach proposed by the author is the first part of the essay.*

<sup>16</sup> «Creating a common understanding about the multiple meanings and significance of authenticity is critical in the digital environment, in which information resources exist in many formats yet are interactive». Cfr. Council on Library and Information Resources, *Authenticity in a digital environment*, Washington D.C., 2000, p. vii.



**Progetti**



# ATHENA: un ponte tra i musei europei ed Europeana

**Maria Teresa Natale**

ICCU

Il progetto europeo ATHENA (Access to cultural heritage networks across Europe), coordinato dall'Istituto centrale per il catalogo unico delle biblioteche (ICCU), è una "rete di buone pratiche" finanziata dal programma *eContentplus*, sviluppato a seguito dei risultati del progetto MINERVA.

Ne fanno parte membri di venti stati dell'Unione Europea e tre osservatori extra-europei, 109 musei di grande rilievo e altre istituzioni culturali sono direttamente associati al progetto fin dal suo inizio.

I suoi obiettivi, da realizzare in 30 mesi (novembre 2008 – aprile 2011), sono:

- sostenere e incoraggiare la partecipazione dei musei e di altre istituzioni culturali a Europeana;
- produrre un set di strumenti, raccomandazioni e linee guida, ponendo l'accento sulle strutture dei dati e sulla loro aggregazione, sulle terminologie e sul multilinguismo, sugli identificatori persistenti, sui diritti di proprietà intellettuale;
- identificare i contenuti digitali presenti nei musei europei;
- contribuire all'integrazione fra i diversi settori del patrimonio culturale con l'obiettivo primario di fondere tutti questi differenti contributi in Europeana, in cooperazione con altri progetti riguardanti più direttamente l'ambito delle biblioteche e degli archivi;
- sviluppare un'infrastruttura tecnologica per consentire l'aggregazione dei metadati e l'interoperabilità semantica con Europeana;
- ampliare il numero delle istituzioni partecipanti.

Scopo di questo contributo è fornire una sintesi dei risultati del progetto dopo due anni dall'avvio<sup>1</sup>.

Nelle fasi iniziali del progetto è stata avviata un'indagine sull'applicazione degli standard di metadati da parte dei musei europei partner del progetto, con l'obiettivo di identificare quelli più rispondenti ai requisiti di Europeana. Le conclusioni<sup>2</sup>

<sup>1</sup> Tutte le informazioni sono disponibili sul sito Web: <http://www.athenaeurope.org>.

<sup>2</sup> Gordon McKenna – Chris De Loof, *Report on existing standards applied by European Museums*, D3.1, 30 April 2009, <http://www.athenaeurope.org/getFile.php?id=396>.

hanno evidenziato che nel settore degli standard di metadati culturali vi è un numero molto limitato di standard condivisi e che il cammino verso l'interoperabilità è ancora molto lungo. In molti paesi, come in Italia, vengono adottati standard nazionali, mentre, per quanto riguarda quelli tecnici, ne vengono utilizzati di specifici per ogni tipo di risorsa multimediale, spesso a sorgente aperta. Non è invece chiaro lo scenario di utilizzo degli standard relativi alla gestione dei diritti di proprietà intellettuale, soprattutto nelle piccole istituzioni. A seguito di quest'indagine è stata prodotta una pubblicazione molto agile che cataloga e descrive sinteticamente gli standard di metadati che i musei, gli archivi e le biblioteche europee dovrebbero conoscere quando intendono avviare un'attività di digitalizzazione<sup>3</sup>.

Tra gli obiettivi prioritari del progetto c'è anche l'identificazione delle raccomandazioni e delle buone pratiche sull'applicazione degli standard e dei formati di harvesting dei metadati. Se per gli standard tecnici si è suggerito di seguire, con poche modifiche e aggiunte, le linee guida elaborate nell'ambito del progetto MINERVA (Ministerial NETwork for Valorising Activities in Digitisation)<sup>4</sup>, riguardo a quelli sui metadati si sono analizzati i tre diversi ambiti del loro utilizzo: la gestione delle raccolte, l'erogazione di servizi, la ricerca online. Lo studio<sup>5</sup> ha evidenziato come i metadati utilizzati nella gestione delle raccolte fossero molto ricchi e come, se opportunamente soggetti ad harvesting, avrebbero potuto svolgere un ruolo chiave nel fornire un buon servizio informativo agli utenti. Si poneva quindi il dilemma di quale formato di harvesting dei metadati adottare nell'ambito del progetto ATHENA, considerando che – secondo l'opinione unanime degli esperti dei musei – gli schemi DC (Dublin Core) ed ESE (Europeana Semantic Elements) non sono sufficienti a illustrare adeguatamente i contenuti museali. Queste considerazioni hanno spinto il progetto ATHENA ad adottare LIDO (Light Information Describing Objects), ritenuto il formato più adatto per attuare l'harvesting dei dati museali in una piattaforma di servizio.

LIDO<sup>6</sup> quindi non è un nuovo schema di metadati, ma per l'appunto un formato di harvesting elaborato da esperti a livello internazionale sulla base di standard e buone pratiche applicate in numerosi paesi non solo europei, come CDWA Lite (Categories for the Description of Works of Art Lite), CIDOC-CRM (CIDOC Conceptual Reference Model), Museumdat e SPECTRUM.

<sup>3</sup> Gordon McKenna – Chris De Loof, *Digitisation: standards landscape for European museums, archives, libraries*, Roma, 2009, <http://www.athenaeurope.org/getFile.php?id=435>.

<sup>4</sup> MINERVA Technical guidelines for digital cultural content creation programmes: Version 2.0, 2008, editors: Kate Fernie, Giuliana De Francesco, David Danson, <http://www.minervaeurope.org/interoperability/technicalguidelines.htm>.

<sup>5</sup> Gordon McKenna – Chris De Loof, *Recommendations and best practice report regarding the application of standards, including recommendations for a harvesting format and fact sheets for dissemination*, D3.2, 31 July 2009, <http://www.athenaeurope.org/getFile.php?id=538>.

<sup>6</sup> Per un approfondimento su LIDO, si rimanda all'articolo di Regine Stein, in questo stesso fascicolo, p. 157-160

Identificato quindi lo standard comune per l'harvesting dei metadati, il partner tecnologico del progetto ha proceduto allo sviluppo di una piattaforma che fungesse da "aggregatore" dei metadati forniti dalle istituzioni partecipanti con l'obiettivo di rendere disponibili i metadati su Europeana.

Questo Web service, tuttora in corso di sviluppo, è stato realizzato con l'integrazione di tecnologie open source ed è caratterizzato da un'interfaccia amichevole e piuttosto intuitiva. L'applicazione consente di:

- registrare l'istituzione (*content provider*) all'interno del sistema;
- mappare i metadati della singola istituzione sullo schema di harvesting LIDO;
- caricare i propri set di metadati (*ingestion*);
- pubblicarli nel repository di progetto.

Il sistema consente anche di monitorare lo stato dell'*ingestion* dei fornitori di contenuti e di effettuare analisi statistiche.

I dati caricati dalle singole istituzioni, dopo un opportuno controllo di qualità, vengono raccolti mediante harvesting dal server di ATHENA nella piattaforma tecnologica di Europeana, attraverso il protocollo OAI-PMH.

La conseguenza di questa procedura è che i dati presenti sul server risultano arricchiti rispetto ai dati trasferiti a Europeana e mappati sul profilo applicativo ESE, attualmente adottato presso il portale europeo, che a breve sarà sostituito dal nuovo EDM (Europeana Data Model)<sup>7</sup>.

Allo stato attuale ATHENA ha già reso disponibili su Europeana due milioni di metadati afferenti a molteplici istituzioni culturali europee. Ciò è stato possibile grazie a un'articolata attività di coordinamento che ha:

- definito le specifiche dell'*ingestion plan* e istituito una rete di referenti nazionali per la gestione dei rapporti con i singoli fornitori di contenuti in tutti i paesi partner di progetto;
- elaborato materiale informativo di supporto e organizzato seminari periodici di training sulla mappatura verso LIDO e sull'utilizzo della piattaforma tecnologica;
- attivato un help-desk di supporto ai fornitori di contenuti;
- identificato le procedure per il controllo di qualità, il monitoraggio delle attività, il dialogo tra tutti gli stakeholders coinvolti (fornitori di contenuti, staff tecnico-scientifico, Europeana Ingestion team);
- curato l'adesione di nuovi fornitori di contenuti non inclusi fin dall'inizio tra i partner di progetto.

<sup>7</sup> *Europeana Data Model Primer*, edited by Antoine Isaac, 5 August 2010, [http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605](http://version1.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605).

Tutta quest'attività, inizialmente sottovalutata, si è rivelata fondamentale per il successo del progetto, tenuto conto della complessità delle procedure di *ingestion* applicate a diversi scenari:

- i fornitori di contenuti effettuano essi stessi le procedure di *ingestion* dei metadati nella piattaforma tecnologica di ATHENA;
- l'*ingestion* dei metadati viene effettuata dal punto di contatto nazionale che effettua la procedura per conto di diversi fornitori di contenuti;
- il fornitore di contenuti deve rendere disponibili i propri metadati nell'aggregatore nazionale (è il caso dell'Italia, dove l'aggregatore nazionale è rappresentato da CulturalItalia), che a sua volta funge da fornitore di contenuti di Europeana (in questo caso è stata effettuata una mappatura tra LIDO e il Pico Application Profile).

Grazie all'esperienza nell'aggregazione dei contenuti, ATHENA ha anche coordinato, insieme ad Europeana, due indagini sullo stato dell'arte degli aggregatori nazionali esistenti, realizzate allo scopo di definire delle linee guida per l'installazione di aggregatori per fornire contenuti a Europeana<sup>8</sup>.

Uno specifico gruppo all'interno del progetto sta lavorando nel settore delle terminologie (lessici, dizionari, folksonomie, glossari, classificazioni, thesauri, vocabolari controllati, ontologie, ecc.). Inoltre un'indagine ha consentito di elencare i lessici in uso presso i musei, alcuni dei quali sono multilingue o seguono il linguaggio formale SKOS (Simple Knowledge Organization System). In seguito il gruppo ha lavorato all'elaborazione di una serie di linee guida che favoriscono l'uso di thesauri al fine di permettere la focalizzazione terminologica su specifici ambiti con la creazione di relativi "ponti", nonché di sostenere l'orientamento a favore dell'utente generico. Queste indicazioni sono particolarmente indirizzate alle istituzioni che vogliono rendere visibili i metadati delle proprie collezioni su Europeana. Un ulteriore obiettivo è la sperimentazione di un thesaurus multilingue su un ambito specifico e con un numero limitato di voci per verificare la sua efficacia sul portale europeo e la correttezza delle linee guida proposte<sup>9</sup>.

Un'ulteriore attività portata avanti dal progetto ATHENA è quella collegata ai diritti di proprietà intellettuale. Un gruppo di lavoro ristretto ha contribuito, assieme al coordinamento del progetto, allo sviluppo della licenza di Europeana, rappresentando gli interessi dei partner di ATHENA.

<sup>8</sup> Europeana, *Europeana Aggregators' Handbook*, 1 May 2010, [http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=94bcddbf-3625-4e6d-8135-c7375d6bbc62&groupId=10602](http://version1.europeana.eu/c/document_library/get_file?uuid=94bcddbf-3625-4e6d-8135-c7375d6bbc62&groupId=10602).

<sup>9</sup> Le attività e i risultati di questo gruppo di lavoro europeo sono disponibili sul Wiki del progetto: <http://www.athenaeurope.org/athenawiki/>.

Il gruppo di lavoro ha inoltre realizzato uno studio sulla legislazione relativa ai diritti di proprietà intellettuale in relazione agli obiettivi di Europeana<sup>10</sup>, tenendo conto dei maggiori trattati internazionali, delle direttive comunitarie e delle legislazioni nazionali in relazione agli istituti culturali.

Questo stesso gruppo sta lavorando all'elaborazione di una *Step-by-step guide* sugli *IPR issues*, un vero e proprio servizio online che possa supportare i fornitori di contenuti culturali nel determinare lo stato dei diritti di proprietà intellettuale relativi ai materiali che si intendono divulgare.

<sup>10</sup> Barbara Dierickx, *Overview of IPR legislation in relation to the objectives of Europeana: 1 November 2008 – 30 April 2009*, in collaboration with Rony Vissers, D6.1, 31 July 2009, <http://www.athenaeurope.org/getFile.php?id=335>.

# Il trattamento delle immagini digitali alla Biblioteca estense universitaria di Modena (1990-2010)

**Andrea Palazzi**

*Biblioteca estense universitaria di Modena*

L'esperienza della Biblioteca estense universitaria (BEU) nel campo delle immagini digitali risale ai primi anni Novanta. Nel corso di questi vent'anni l'attività si è andata precisando sia come ricerca di soluzioni razionali ed efficienti in grado di rispondere alle crescenti richieste del pubblico, sia come individuazione di soluzioni innovative di digitalizzazione per tutelare e valorizzare il patrimonio della biblioteca. Ai fini di un'esposizione possibilmente completa e al tempo stesso sintetica e non dispersiva, sembra utile scandire il discorso in sezioni, limitando allo stretto necessario gli aspetti tecnici di ogni soluzione applicata.

## Leggio elettronico

Il progetto Leggio elettronico risale agli anni 1991-1995, cioè agli albori dell'era Windows, quando la grande maggioranza dei pc dell'amministrazione pubblica lavorava ancora in modalità testuale e non grafica. Una selezione di manoscritti e mappe di grande pregio fu fotografata e le diapositive vennero digitalizzate. Il progetto prevedeva

«per ciascuno dei codici trattati, un primo trattamento sistematico dei volumi (comprensivo di eventuali carte di guardia, carte di solo testo, ecc.) secondo una metafora di leggio elettronico. Il sistema digitale consentirà in tal modo un utilizzo del sistema di documentazione in qualità di sostituto degli originali per la maggior parte delle esigenze di consultazione e di studio, nell'ambito delle possibilità di un mezzo a uscita su monitor e pertanto più congeniale alla visualizzazione di immagini a colori che non alla lettura di testi»<sup>1</sup>.

Come si può osservare, fin dall'inizio il trattamento delle immagini digitali in Estense diede priorità agli aspetti connessi con la tutela degli originali, analizzando differenti modalità di fruizione.

<sup>1</sup> Attività culturali e tecnologie avanzate SpA (ACTA), *Offerta per il trattamento informatizzato mediante ns. sistema digitale di immagini di codici miniati*, Firenze 14 febbraio 1992, Archivio della Biblioteca estense universitaria di Modena, 569-II/1992.

Questo primo progetto dovette però misurarsi con le limitazioni hardware e software dell'epoca: in quel periodo la compressione delle immagini era ancora poco nota, e per le immagini non compresse, in ambiente Windows 3, era usuale il ricorso al formato BMP (bitmap). Fu adottata una soluzione di compressione attraverso scheda hardware che, se velocizzava decisamente la gestione delle immagini sui computer del periodo (80386 e 80486, i Pentium sarebbero apparsi solo con Windows95), non ricorreva però a tecnologie standard, sia pure appena nate, come la compressione JPEG. Era inoltre ancora da mettere a punto la fedeltà cromatica all'originale e la capacità di dettaglio era fortemente limitata dalla bassa risoluzione delle immagini (il riferimento di quegli anni era: 1024x768 pixel con risoluzione a 72 DPI, dots per inch).

Allestita una postazione nella Sala mostra dell'Estense, la soluzione del leggio riscosse tuttavia un discreto interesse, ma forse più per la novità e l'attenzione che l'amministrazione stava riservando al nascente mondo digitale che per un'effettiva fruizione quotidiana. Oltre ai limiti hardware, si poté constatare come non fosse affatto intuitivo per il pubblico passare dalla consultazione analogica a quella virtuale, tanto più in una fase in cui anche negli ambienti grafici era ignota qualunque standardizzazione: il significato di ogni icona andava indagato e scoperto, rendendo l'interfaccia nel complesso assai poco amichevole.

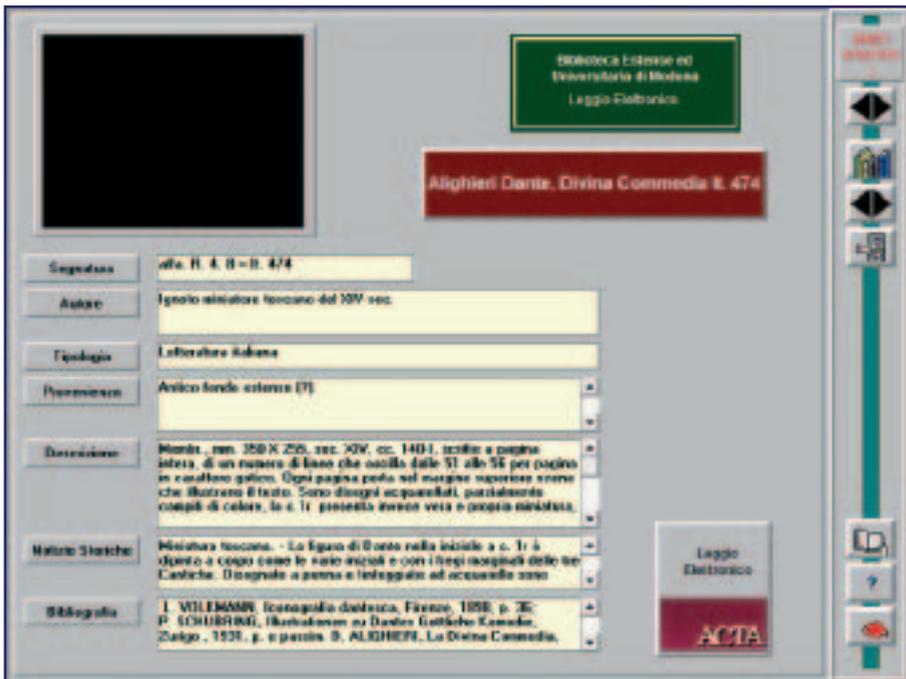


Figura 1. Leggio elettronica, ricerca mediante il database

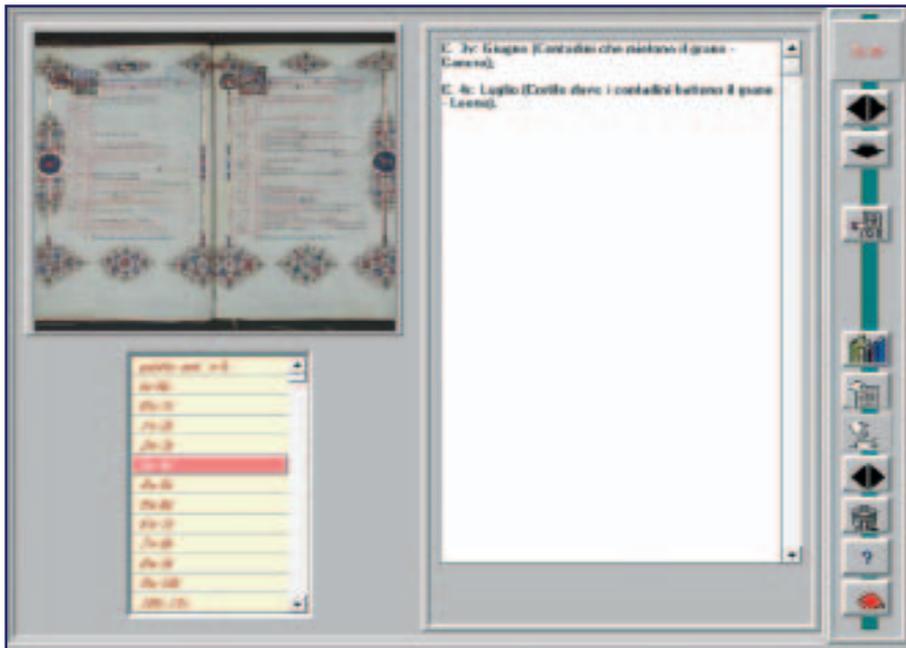


Figura 2. Leggio elettronico, visualizzazione delle carte



Figura 3. Leggio elettronico: ingrandimento e particolari disponibili con ulteriore ingrandimenti

## Carta degli stati estensi

Di tali limiti, tecnici e progettuali, si tenne conto nel 1998-2001 per il successivo progetto, realizzato in collaborazione con la Fotoscintifica di Parma. Si trattava in questo caso della digitalizzazione della *Carta degli stati estensi*<sup>2</sup> di Marco Antonio Pasi, risalente al 1580, cioè a subito prima della devoluzione di Ferrara alla Santa Sede e dello spostamento della capitale a Modena. L'esemplare estense è una copia, con varianti, di quello conservato nell'Archivio di Stato di Modena.

Val la pena di osservare che nella carta il settentrione è in basso e il levante a sinistra, con rappresentazione ribaltata rispetto a quella ora usuale. La carta presenta un notevole valore non soltanto geografico, ma anche socio-economico (vi appare la *grande bonificazione ferrarese* del secondo Cinquecento, pressoché contemporanea al documento), demografico (i centri abitati sono descritti con una certa precisione) e, non ultimo, come rappresentazione dello *status* della corte ferrarese, come mostra l'evidenziazione delle famose *delizie estensi*, luoghi di svago della corte<sup>3</sup>.

Anche solo per le dimensioni, le carte sono di difficile conservazione: la carta dell'Estense misura 312x276 cm (divisa attualmente in otto fogli telati, ciascuno di circa 79x87 cm). Poiché, terminata la funzione per cui era stata creata, la carta venne conservata in condizioni precarie, il progetto ne prevedeva il restauro virtuale, così da restituire la leggibilità originaria senza dover intervenire sull'originale, se non allo scopo di fermarne il deterioramento, ma conservando i sedimenti delle sue vicende storiche.

Il progetto prevedeva sin dall'inizio di fotografare le carte direttamente in formato numerico, ricorrendo a un dorso digitale. Per garantire una buona risoluzione (300 DPI) ogni foglio venne virtualmente suddiviso in 9 riquadri, che vennero poi riasssemblati in post-processing.

Inizialmente era prevista solo una riproduzione dinamica della carta che, mediante il software Apple Quicktime, avrebbe consentito di spostarsi all'interno dei fogli e tra i fogli, permettendo un ottimo livello di dettaglio tramite zoom. L'Estense chiese esplicitamente che oltre al formato Quicktime venissero consegnati, in formato TIFF, anche i 72 riquadri ripresi prima del restauro virtuale e gli 8 fogli rimontati e restaurati digitalmente. Queste immagini fanno tuttora parte della banca dati iconografica della biblioteca e vi si è fatto ricorso varie volte per soddisfare richieste degli utenti. Dei fogli interi restaurati si sono poi realizzate stampe in scala 1:1, che vengono fornite in consultazione invece degli originali; questi ultimi restano naturalmente disponibili per esigenze di studio particolari.

<sup>2</sup> La *Carta degli stati estensi* è consultabile, in versione ridotta, sul sito della Biblioteca estense, all'indirizzo: <http://www.cedoc.mo.it/estense/img/geo/StatiEstensi/index.html>.

<sup>3</sup> Per una più approfondita analisi della carta della biblioteca, si rinvia al saggio di Laura Federzoni, *La Carta degli stati estensi di Marco Antonio Pasi: Il ritratto dell'utopia*, in *Alla scoperta del mondo: L'arte della cartografia da Tolomeo a Mercatore*, Modena: Il Bulino, 2001; mentre per l'esemplare dell'archivio si può consultare lo scritto di Alessandra Chiappini, *Il territorio ferrarese nella carta inedita dei Ducati Estensi di Marco Antonio Pasi (1571)*, «Atti e memorie della Deputazione provinciale ferrarese di storia patria», serie 3, 1973, vol. 13.

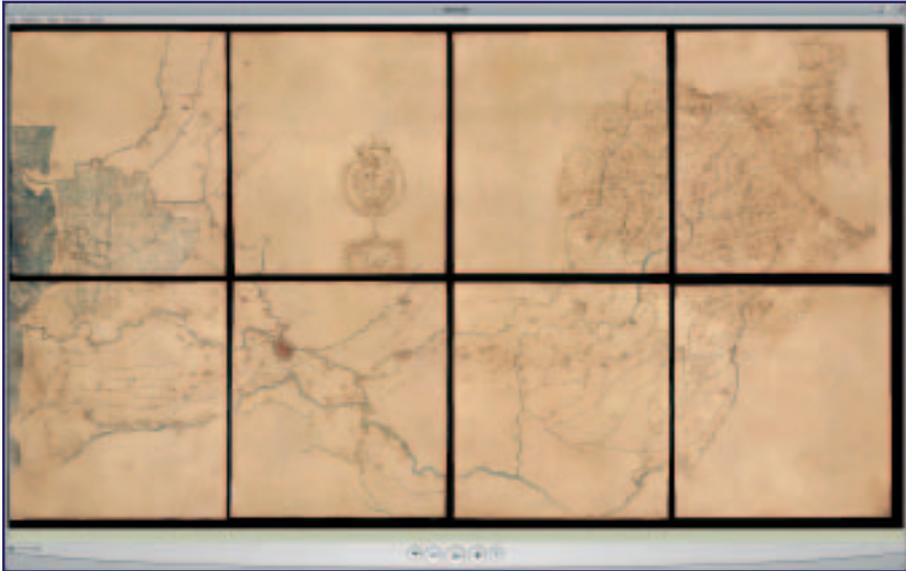


Figura 4. Carta degli stati estensi: visione di insieme nella navigazione con Quicktime

È da notare che ogni file dei fogli restaurati si aggira sui 350-380 megabyte: dimensioni di difficile gestione con i pc di allora, molto più maneggevoli con quelli odierni. Lo stesso vale per il formato Quicktime, in cui i fogli restaurati sono stati compressi, fino a giungere a un file circa di 160 megabyte, dimensioni abbastanza usuali al giorno d'oggi per questo tipo di dati.



Figura 5. Carta degli stati estensi: Modena e dintorni nella navigazione con Quicktime



Figura 6. Carta degli stati estensi: zoom su Modena nella navigazione con Quicktime



Figura 7. Carta degli stati estensi: particolare di Modena nelle condizioni reali



Figura 8. Carta degli stati estensi: particolare di Modena dopo il restauro virtuale

### Archivio muratoriano

A partire dal 1999, e quasi contemporaneamente all'esperienza della *Carta*, l'Estense scelse di ricorrere alle immagini digitali per un progetto di salvaguardia ad ampio respiro, la riproduzione e il restauro dell'intero archivio di Lodovico Antonio Muratori (1672-1750), bibliotecario e archivista ducale, considerato il fondatore della storiografia moderna.

L'Archivio muratoriano costituisce uno dei principali fondi dell'Estense, e – come molti altri archivi di letterati e studiosi – è ordinato per filze e fascicoli e diviso in due grandi sezioni. La prima raccoglie i manoscritti delle opere erudite (a volte con lettere di accompagnamento o relative al contenuto dell'opera), quaderni giovanili, diplomi accademici e materiali vari; la seconda divide il carteggio in fascicoli – ordinati secondo il nome dei 2.054 corrispondenti – che contengono gli originali in arrivo, minute o copie delle lettere di Muratori ed eventuali allegati, spesso schizzi epigrafici.

Una ricognizione preliminare individuò circa 300.000 immagini da riprodurre, di formato generalmente di poco inferiore all'A4. In TIFF non compresso a 300 DPI, circa 20 megabyte a immagine, avrebbero prodotto 6 terabyte complessivi, valore

che nel 2000 era ben lungi dall'essere gestibile, e che si giudicò non indispensabile perché, trattandosi di un fondo archivistico, prevalsero esigenze di semplice leggibilità rispetto alla fedeltà richiesta dai codici miniati. Sulla base delle richieste di riproduzione che man mano pervenivano, e in considerazione delle possibilità di gestire concretamente la base dati iconografica che si sarebbe formata nel tempo per salvaguardare gli originali, si decise di ricorrere a macchine fotografiche e non a scanner, con immagini sempre a 300 DPI ma in formato JPEG poco compresse, per un peso di circa 3-4 megabyte a immagine. Le immagini vennero consegnate su supporti ottici, prima CD e poi DVD, e riversate sui server dell'Estense.

Dopo due cospicui stanziamenti iniziali del Ministero il progetto venne purtroppo rallentato dalla scarsità di fondi. Attualmente la base dati comprende più di 126.000 immagini, per un'occupazione complessiva di oltre 435 gigabyte.

Sin dall'inizio si pose naturalmente il problema del reperimento e della consultabilità delle immagini. L'Archivio viene tuttora consultato ricorrendo al catalogo-inventario ottocentesco di Luigi Vischi<sup>4</sup>, risalente a prima che l'archivio entrasse in Estense, e una copia del quale, ora disponibile in PDF sul sito della biblioteca<sup>5</sup>, è stata nel corso del Novecento annotata e corretta dai bibliotecari estensi.

Si immise quindi il catalogo nella tabella di un database (formato Microsoft Access); in un'altra tabella venne importato l'indice dei nomi del catalogo (realizzato nel 1983 e in seguito immesso a computer in formato testuale, ora disponibile sul sito insieme al catalogo); una terza tabella elencava percorso e nome del file di ogni JPEG realizzato. Mettendo in connessione queste tabelle si realizzò un prototipo funzionante di RDBMS (Relational database management system) dove poter cercare le immagini a partire dal nome dei corrispondenti, oppure scorrendo sequenzialmente i fascicoli virtuali.

Come si vedrà oltre, attualmente le immagini sono disponibili sulla LAN (Local Area Network) della biblioteca e vengono consultate quotidianamente in luogo degli originali. Quanto alla loro disponibilità sull'internet, va ricordato che il progetto nacque a scopo essenzialmente conservativo e ben prima della diffusione e della standardizzazione dei metadati, in particolar modo dei MAG (metadati amministrativi gestionali). Con la consulenza dell'Istituto centrale per il catalogo unico (ICCU), si è quindi deciso di munire le immagini già realizzate di metadati che consentano di identificarle singolarmente e univocamente. Poiché l'inventario di Vischi presenta spesso alternanze formali e soprattutto non individua i singoli documenti, questo passaggio implica la catalogazione del fondo secondo gli standard correnti, effettuata la quale si potrà continuare il lavoro di ripresa e incrementare la base iconografica.

<sup>4</sup> *Archivio Muratoriano preceduto da una lettera inedita di Lodovico Ant. Muratori intorno al metodo de' suoi studi, per cura di L. V. [Luigi Vischi]*, Modena: Zanichelli, 1872.

<sup>5</sup> <http://www.cedoc.mo.it/estense/info/cataloghi-mss.html>.

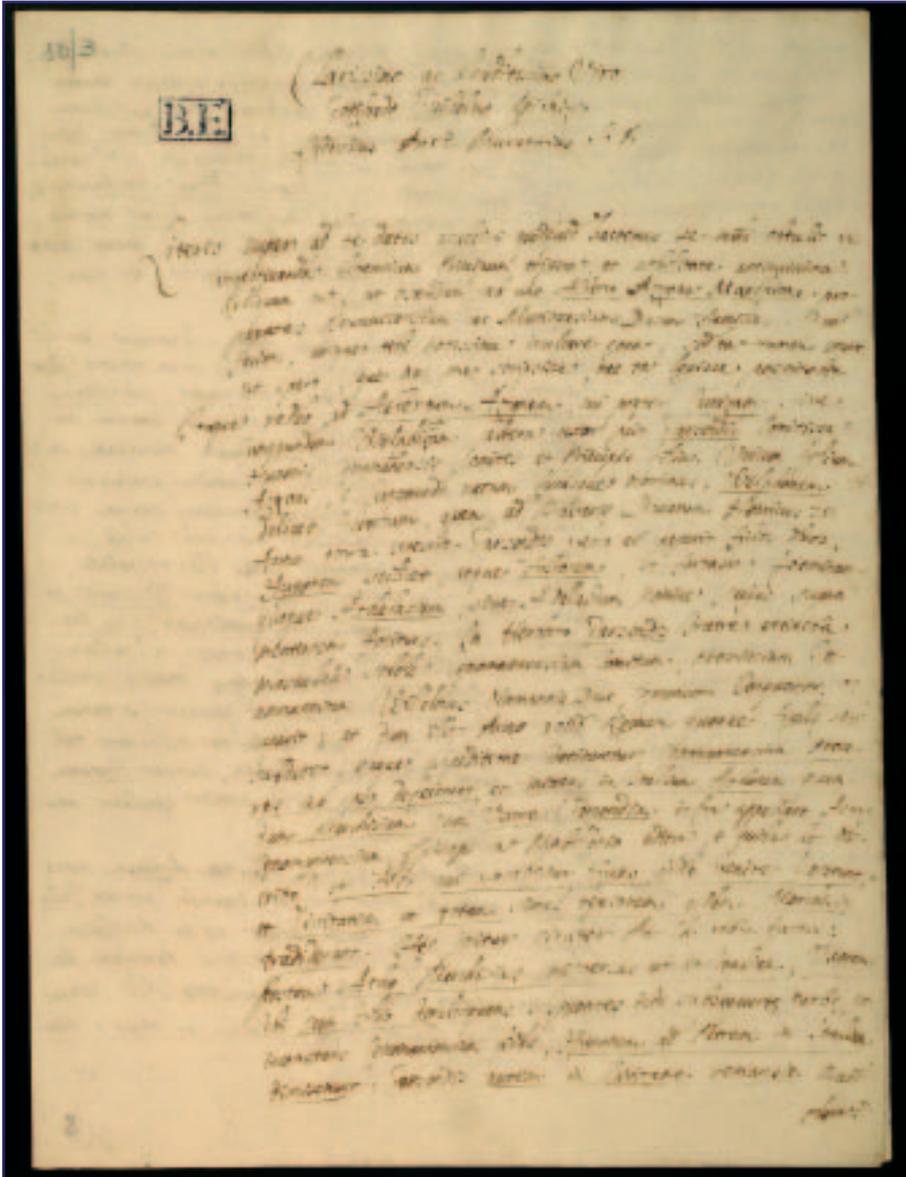


Figura 9. Archivio muratoriano: note di G.W. v. Leibnitz sulle origini della famiglia estense

### Catalogo Staderini (Progetto BDI – Cataloghi storici)

Contemporaneamente alla digitalizzazione dell'Archivio muratoriano, l'Estense aderì al progetto BDI – Cataloghi storici concentrandosi sul catalogo Staderini (oltre 250.000 schede), che si riferisce alla maggioranza delle opere pervenute in biblioteca dalla metà dell'Ottocento sino al 1958.

Il progetto venne attuato secondo gli standard richiesti, con una sola variante: il campo di testo che trascriveva la scheda venne portato dai 20-40 caratteri previsti inizialmente a circa 100. Vicissitudini di vario genere ne complicarono l'attuazione, il che ha implicato negli anni 2005-2009 una radicale revisione delle trascrizioni curata dai bibliotecari dell'Estense.

Alla fine del 2009 si sono potuti mandare all'ICCU dati e immagini per la pubblicazione sul portale della Biblioteca digitale italiana<sup>6</sup>, mentre in LAN si è scelta una soluzione leggermente diversa. Le immagini, originariamente in formato TIFF CCITT Fax4 (b/n, 1 bit/pixel), come da parametri del progetto, sono state convertite in formato PNG (Portable Network Graphics), che richiede alcuni byte in più ma viene in compenso riconosciuto nativamente da tutti i browser disponibili ed è di fatto uno degli standard per le immagini *lossless* (senza perdita di informazioni, a differenza del JPEG). I dati sono stati riversati in un database Microsoft SQL Server, che viene interrogato in ASP (Active Server Pages). Quindi tutti i client della LAN possono effettuare ricerche, sfruttando gli operatori booleani, e visualizzare immediatamente le schede nel browser, senza bisogno di ricorrere a installazioni sui singoli computer o a plugin particolari (di implementazione complessa in una LAN dalle policies di sicurezza rigide come quella estense); visualizzando il PNG, il browser consente inoltre lo zoom in/out con un semplice clic sull'immagine. Dato che viene effettuata una ricerca a testo libero, i risultati sono cospicui, bastino questi esempi (i risultati sono quantificati in pagine HTML che elencano 10 schede ognuna):

<i>stringa cercata</i>	<i>operatore</i>	<i>risultati</i>
modena ; modona ; mutina	OR	467
modena	-	455
modena ; bologna	OR	541
modena ; bologna	AND	3
modena ; bologna	NOT	452 (correttamente, 455 -3)

Sempre con l'interfaccia del browser, è possibile cercare la sola parte iniziale della scheda, che nel vecchio catalogo corrisponde in genere all'intestazione:

<i>stringa cercata</i>	<i>isultati</i>
muratori (testo libero)	87
muratori (parte iniziale)	35

<sup>6</sup> <http://catalogohistorici.bdi.sbn.it/code/index.asp>.



Figura 10. Catalogo Staderini: ricerca con NOT



Figura 11. Catalogo Staderini: ricerca con OR

## Teca digitale

Parallelamente alle iniziative finora ricordate, a partire dagli inizi degli anni Duemila si sono intensificate le richieste da parte degli utenti di riproduzioni digitali di opere conservate in Estense, soprattutto manoscritti, ma anche opere antiche a stampa e talora periodici. Sino al 2009 le riprese venivano effettuate da uno studio fotografico esterno. Dal 2010, grazie alla collaborazione dell'Archivio di stato di Modena, la biblioteca può ricorrere per due giorni la settimana a un fotografo in organico all'amministrazione, insieme al quale si sono studiate le migliori modalità di ripresa nei limiti concessi dalle risorse disponibili.

Quando le riproduzioni riguardano un'intera opera o un intero fascicolo, manoscritto o a stampa, la biblioteca archivia una copia delle immagini. A queste si uniscono quelle ricevute gratuitamente dalla Fondazione di Vignola<sup>7</sup> sulla base di un'apposita convenzione, e quelle realizzate direttamente dalla biblioteca, da originale o da facsimile, in occasione di mostre ecc., senza considerare le copie ricevute dagli editori nel caso di riproduzioni destinate a facsimili, naturalmente di ottima qualità (nel caso del manoscritto più prezioso della biblioteca, la celebre *Bibbia* di Borso d'Este, si sono ricevute circa 600 diacolor 20x25, che si sono fatte scandire in service a 1000 DPI; se ne sono poi ricavate le immagini delle singole pagine, di circa 80 megabyte l'una, poiché la ripresa per il facsimile era stata effettuata a doppia carta con il manoscritto sfasciolato, e le pagine naturalmente non risultavano sequenziali).

Lo standard di riferimento per le acquisizioni è 300 DPI (su richiesta del Ministero, sono state a volte ricavate immagini a 600 DPI, raramente richieste dagli utenti, anche a scopo tipografico; per alcuni facsimili la biblioteca ha inoltre ricevuto immagini a 400 DPI); gli interventi software effettuati dalla biblioteca si limitano di norma alla rifilatura delle immagini: si fa ricorso al software Adobe Photoshop CS4 solo per equilibrare riproduzioni dal cromatismo complesso, per produrre in batch la copia a 150 DPI delle immagini destinate alla teca pubblica in LAN, e a volte per rendere le immagini meglio visibili per le mostre e sul sito (dove vengono ridotte a 400 pixel), nel qual caso si lavora naturalmente su copie dei file archiviati, che restano inalterati. Per riprodurre gli originali si ricorre di norma a una macchina fotografica (attualmente una Canon EOS 5D Mk II da 21 megapixel, con obiettivo 24-70 mm e diaframma 2.8); per i facsimili, per le fotografie e per altro materiale non legato, quando lo consentono le condizioni di conservazione, a uno scanner piano (attualmente un Kodak i1420 con lastra piana, che scandisce con ottima fedeltà cromatica un A3 a 300 DPI in meno di 5" e può arrivare a 1200 DPI ottici, usati a volte per alcuni dettagli, specie dalle fotografie in occasione di mostre). Dallo scanner si ottengono file TIFF; dalla macchina fotografica file JPEG e CR2 (l'attuale implementazione Canon del formato RAW, del quale ogni produttore di macchi-

<sup>7</sup> [http://www.cedoc.mo.it/estense/info/collab/2008\\_vignola/index.html](http://www.cedoc.mo.it/estense/info/collab/2008_vignola/index.html).

ne digitali ha una sua versione proprietaria, e per il quale Adobe sta cercando di promuovere uno standard comune, il DNG, Digital Negative).

Le immagini vengono quindi ordinate in cartelle – sulla base della collocazione, unico elemento presente sia nei manoscritti sia nelle opere a stampa – su uno storage server (in mirroring con un server gemello). Per distinguere varie riproduzioni della stessa opera il nome delle cartelle alla segnatura fa seguire un codice progressivo e il numero di DPI. I relativi dati bibliografici e fotografici (hardware, software, DPI, ecc.) vengono immagazzinati in un database, con interfaccia stesa in Microsoft Visual Basic 6, che ne genera un listato in formato HTML di questo tipo:

<b>ALFA.C.2.32</b>	ID: 21 Opera: Stultifera navis Sec.: 15. Cartella Teca: alfa.c\alfa.c.2.32_[001-150](stultifera-navis) Riproduzione completa/parziale/tav.: Tav. Immagini: 118 - Qualità (1-5): 3 - DPI: 150 - MB: 77 Da originale/facsimile: O - Foto/scansione: S Esecuzione: BEU - Anno: 2004 hardware: Microtek 9800 - software: Adobe PS 7
<b>ALFA.D.2.24</b>	ID: 80 Opera: Isagogae breves Cartella Teca: alfa.d\alfa.d.2.24_[001-150] Riproduzione completa/parziale/tav.: P Immagini: 13 - Qualità (1-5): 3 - DPI: 150 - MB: 11 Da originale/facsimile: O - Foto/scansione: F Esecuzione: Negro, V. - Anno: 2008 hardware: Canon EOS 5D - software: Adobe PS CS 1
<b>ALFA.D.5.22</b>	ID: 16 Opera: Apocalypsis Iohannis Cartella Teca: alfa.d\alfa.d.5.22_[001-150](foto-orig) Riproduzione completa/parziale/tav.: C Immagini: 108 - Qualità (1-5): 3 - DPI: 150 - MB: 58 Da originale/facsimile: O - Foto/scansione: F Esecuzione: Negro, V. - Anno: 2008 hardware: Canon EOS 5D - software: Adobe PS CS 1 * Realizzato con il contributo della Fondazione di Vignola
<b>ALFA.D.5.22</b>	ID: 15 Opera: Apocalypsis Iohannis Cartella Teca: alfa.d\alfa.d.5.22_[002-150](scan-fax) Riproduzione completa/parziale/tav.: C Immagini: 108 - Qualità (1-5): 3 - DPI: 150 - MB: 61 Da originale/facsimile: F - Foto/scansione: S Esecuzione: BEU - Anno: 2009 hardware: Kodak i1420 - software: Adobe PS CS 4

Questo indice è disponibile sui pc destinati al pubblico e consente di aprire la cartella con un semplice clic sul suo nome, così da scorrere rapidamente le immagini. Come visualizzatore si ricorre al software IrfanView, gratuito per uso non commerciale<sup>8</sup> e di apprendimento intuitivo e immediato, che consente di ingrandire l'immagine, ruotarla, variane luminosità e contrasto, ecc. I cambiamenti applicati dai lettori non possono essere salvati, né si possono, neppure accidentalmente, cancellare le immagini.

Per quanto riguarda l'Archivio muratoriano, per la teca pubblica l'Estense ha scritto un apposito software che ricostruisce virtualmente filze e fascicoli, inserendovi una copia delle immagini, che sono state consegnate alla biblioteca in semplice sequenza numerica per filename (beu\_am\_0000001.jpg, beu\_am\_0000002.jpg, ecc.), per ridurre i margini di errore e semplificare la gestione del database contenente i metadati.

In questo modo si sono potuti ordinare sinora quasi 147.000 file (compresi quelli dell'Archivio muratoriano), per un totale di 1.02 terabyte. Di questi, oltre 140.00 sono già a disposizione del pubblico in LAN e vengono consultati quotidianamente sui pc destinati agli studiosi, due dei quali con monitor 24" a risoluzione 1920x1200.

Altre 198.000 immagini, per 900 gigabyte complessivi, sono in corso di ordinamento e valutazione (non tutte le immagini ricevute si rivelano di qualità accetta-

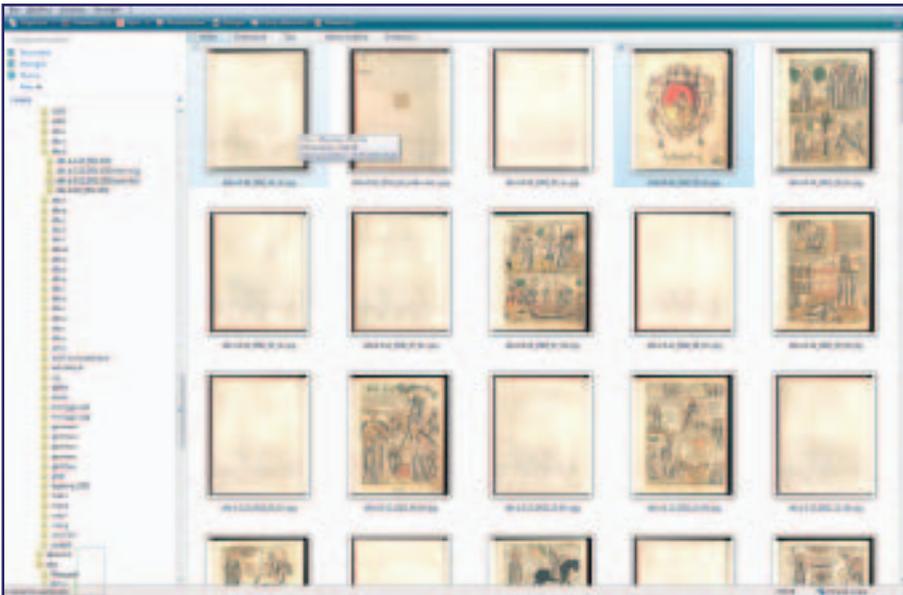


Figura 12. La teca

<sup>8</sup> [http://www.irfanview.com/main\\_what\\_is\\_engl.htm](http://www.irfanview.com/main_what_is_engl.htm).



Figura 13: Visualizzazione delle immagini della teca

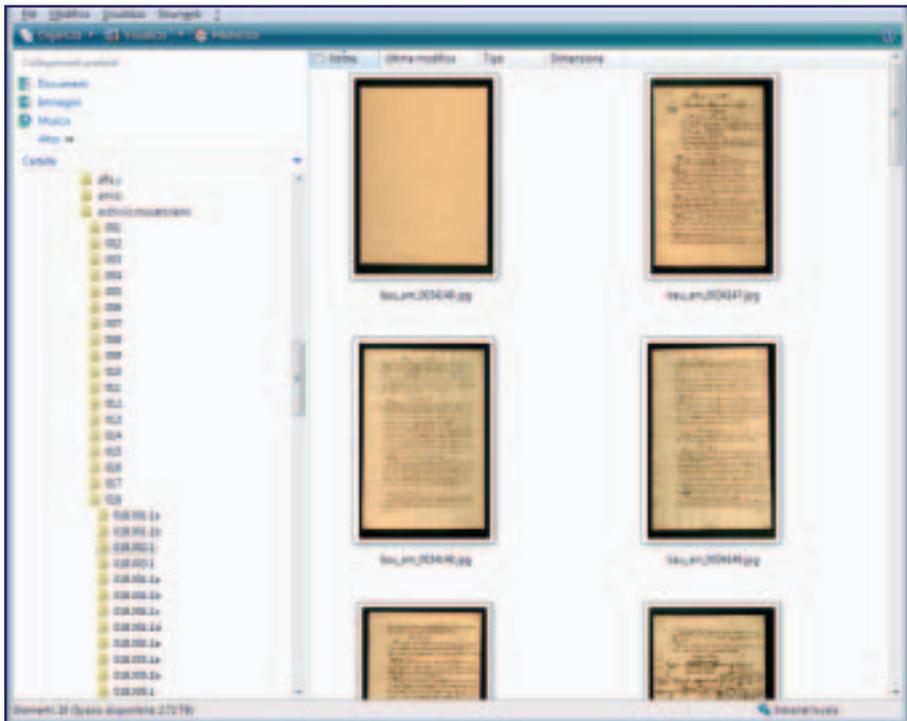


Figura 14: Visualizzazione zoom delle immagini della teca

bile: quelle risalenti agli anni Novanta e quelle consegnate dagli studiosi sono spesso inutilizzabili, ma molte vengono ugualmente archiviate, a scopo di documentazione storica).

A tutte queste immagini ha accesso l'ufficio relazioni con il pubblico dell'Estense, che riscontra le richieste di riproduzione e incarica il fotografo di riprodurre le opere non ancora archiviate. Questa procedura è stata avviata sul finire del 2009, quando cioè si è potuto disporre di storage server capienti in cui concentrare tutto il materiale prodotto, e in questi primi mesi ha già consentito di ridurre sensibilmente i tempi di risposta al pubblico.

### Prospettive attuali e future

Come si è visto, gran parte della base iconografica nasce dall'analisi del workflow lavorativo e dal tentativo di razionalizzarlo e renderlo efficiente con le poche risorse disponibili. Come nel caso dell'Archivio muratoriano, la questione più rilevante resta quella dei metadati: la costituzione di un database relativo alle immagini archiviate nasce proprio dall'idea di corredare dei metadati necessari, in un futuro prossimo, materiali tanto eterogenei. In questa prospettiva va vista l'adesione della biblioteca al progetto World digital library (WDL)<sup>9</sup>, in cui verranno versate le immagini di tre dei più importanti manoscritti estensi. Non tutte le immagini prodotte sinora rispondono comunque ai requisiti minimi richiesti da una pubblicazione formale, né negli anni passati era possibile chiedere di fornire i metadati alle ditte che effettuavano riproduzioni in conto terzi. Lo si è potuto iniziare a fare, specificandolo nel capitolato, in occasione di un set di riproduzioni richiesto dall'Università La Sapienza, ed è ragionevole pensare ad una generalizzazione del metodo.

La biblioteca è attesa inoltre da una grande sfida nel suo immediato futuro. Grazie a un *Protocollo di intesa* stipulato nel novembre 2007 dal Ministero per i beni e le attività culturali, il Comune di Modena e la Fondazione Cassa di risparmio di Modena, il complesso monumentale di Sant'Agostino, utilizzato fino a pochi anni fa come ospedale cittadino, verrà infatti destinato, una volta conclusi i lavori di ristrutturazione il cui termine è previsto per il 2014, ad ospitare un grande polo culturale, al cui interno troverà posto, risolvendo finalmente gli annosi problemi di spazio che tanto ne hanno condizionato l'attività in questi ultimi decenni, anche la Biblioteca Estense. Sulla base di quanto previsto dalle prime ipotesi progettuali formulate dal consorzio recentemente risultato vincitore della gara di progettazione (costituito dal Consorzio Leonardo, dagli architetti Gae Aulenti e Francesco Doglioni) grazie a questa sua nuova collocazione l'Estense potrà allora disporre di circa 5.800 mq, una superficie di due terzi superiore a quella attualmente occupata (pari a circa 3.400 mq).

<sup>9</sup> <http://www.wdl.org/>.

Una simile innovazione, tacendo dell'enorme sfida costituita dal trasferire in altra sede una biblioteca storica della portata e delle dimensioni dell'Estense, non potrà non riflettersi necessariamente sull'organizzazione e sulla concezione stessa dei servizi che la biblioteca potrà offrire ai propri utenti, a partire proprio da quelli legati all'utilizzo delle tecnologie informatiche. Sarà dunque quella un'occasione irripetibile per ripensare nel suo complesso anche l'offerta della biblioteca in tema di immagini digitali online ed offline.

Un ultimo punto riguarda le copie di scorta: come si è accennato, in LAN le immagini sono memorizzate in due NAS (Network Attached Storage) gemelli da 12 TERABYTE l'uno. Per quanto concerne invece il backup remoto, è allo studio una convenzione con l'Istituto per i beni artistici, culturali e naturali della Regione Emilia-Romagna (IBACN), che tra le sue funzioni ha quella di archiviazione e conservazione dei documenti informatici prodotti dalla Regione e, mediante apposite convenzioni, dei documenti prodotti da Province, Comuni e altri soggetti pubblici.

# Imago et umbra

## Programma di digitalizzazione per l'Archivio storico della Pontificia Università Gregoriana: criteri, metodi e strumenti

**Andrea Bozzi – Martín Maria Morales – Marco Rufino**

*Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa – Archivio storico della Pontificia Università Gregoriana – Fondazione rinascimento digitale*

### Criteri<sup>1</sup>

Il titolo di questo progetto fa riferimento a un'espressione che si trova all'interno del *corpus* degli scritti di Nicolò Cusano («Sermo LXI, [...] Vita igitur istius mundi non est vita, sed imago et umbra vitae verae; ita de sapientia et prudentia et laetitia et singulis aliis») con la quale vengono indicate le realtà di questo mondo che, sebbene da una parte rappresentano (*imago*) nel contempo non lo fanno in modo totale e assoluto (*umbra*). In maniera analoga il progetto che qui si presenta, pur portando alla luce, in immagini, i documenti, non potrà mai esaurirne i contenuti.

L'Archivio della Pontificia Università Gregoriana (APUG)<sup>2</sup> possiede un patrimonio che testimonia l'attività intellettuale dei professori gesuiti del Collegio Romano, dalla sua fondazione nel 1551 alla soppressione della Compagnia di Gesù ed ancora dal 1824 all'Unità d'Italia. Questo insieme si presenta come esempio unico di raccolta documentaria attraverso la quale è possibile ricostruire la storia della didattica di una delle università tra le più celebri nell'Europa dei secoli XVI-XVII.

I fondi manoscritti oggi conservati presso l'APUG, provenienti dal patrimonio librario del Collegio Romano, sono il risultato di un complesso *iter*, non ancora del tutto concluso, di conservazione, nascondimento e restituzione. Nell'antico Collegio esistevano diverse "librerie", secondo i diversi corsi d'insegnamento: vi era una *bibliotheca* per la teologia, una per la filosofia, una terza per la retorica ed infine una quarta per gli studi inferiori. Oltre a queste quattro ve n'era però anche

<sup>1</sup> Martín Maria Morales (Archivio storico della Pontificia Università Gregoriana).

<sup>2</sup> Per la storia dell'archivio, descrizione del posseduto e i progetti in corso si veda il sito: <http://www.archiviopug.org>.

una quinta costituita da fondi provenienti da importanti lasciti testamentari: la *Bibliotheca maior*, o anche *secreta*<sup>3</sup>. In seguito alla soppressione della Compagnia di Gesù nel 1773 i fondi delle biblioteche paiono essere rimasti inalterati fino al ritorno dei Gesuiti nel 1824, mentre alcuni manoscritti insieme a codici, libri rari ed oggetti preziosi furono murati in un ripostiglio di cui, durante il secolo successivo, si conservò solo un vago ricordo. Fu in seguito all'emanazione della legge di soppressione delle Corporazioni religiose del 19 giugno 1873 che iniziò l'incameramento dei beni delle biblioteche dei diversi ordini. Nella denuncia del 1877 di Bartolomeo Podestà, primo bibliotecario della Biblioteca nazionale centrale di Roma (Bncr), si dà notizia della scoperta del ripostiglio. Tra i documenti rinvenuti, quelli considerati di scarso interesse, furono collocati in una soffitta dello stesso Collegio Romano sede della nascente Bncr. Nel 1948 la biblioteca, trovandosi nella necessità di liberare questa soffitta, decise di restituire il fondo al Generale della Compagnia.

Uno dei progetti intrapresi dall'APUG riguarda l'analisi dello stratificato processo, storico e culturale, che ha portato alla dispersione dei fondi gesuitici così come quelli di altri ordini religiosi. Questa ricerca implica una ricostruzione della composizione originaria dei fondi delle biblioteche del Collegio Romano a partire dai fondi gesuitici oggi conservati presso la Bncr, l'Archivio di Stato di Roma (ASR), la Biblioteca apostolica vaticana (BAV) e l'Archivum romanum societatis Iesu (ARSI). Ovviamente lo strumento principe per avviare questa e ogni altro tipo di ricerca è il catalogo. Dagli anni Quaranta fino al 2008 i ricercatori dell'APUG, per le intricate vicende che hanno interessato l'archivio, sono stati pochi. I cambiamenti di sede, via via più inaccessibile e disagiata, rappresentano, in parte, la caduta dell'interesse per un tipo di materiale che non alimenta più né l'insegnamento né la ricerca. Gli studiosi, qualora fossero riusciti ad accedere all'archivio, incontravano notevoli difficoltà, limitati da strumenti quali un inventario topografico incompleto e inesatto. Solo ad alcuni importanti manoscritti, grazie al contributo di singoli ricercatori, sono stati dedicati studi che, spesso, purtroppo non sono sfociati in pubblicazioni. Ma, come si vedrà, la vita di un archivio dipende anche dalla sua ricerca.

La limitatezza delle risorse economiche e umane per affrontare gli interventi di conservazione e valorizzazione dell'APUG ha condizionato fortemente le scelte operative degli ultimi due anni. Ad ogni modo, il cammino intrapreso si è manifestato vincente. L'impossibilità di acquistare un software per la catalogazione ha permesso:

- di avviare una collaborazione con l'Istituto centrale per il catalogo unico (ICCU)

<sup>3</sup> Cfr. Costanzo Bizzocchi, *La Biblioteca Segreta del Collegio Romano*, «Gesuiti della Provincia Romana», n. 3, 1971, p. 17.

- e l'acquisizione di MANUS integrando così l'APUG nel progetto di censimento nazionale dei manoscritti;
- di intraprendere un proficuo rapporto con la Fondazione rinascimento digitale di Firenze e con l'Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa (ILC) che consente l'elaborazione di un modulo specifico di Pinakes (d'ora in avanti citato con la sigla PK<sup>4</sup>) e Pinakes Text (PKT) per la realizzazione di un archivio digitale.

Inoltre, l'esiguità dei fondi investibili per la realizzazione di un laboratorio di digitalizzazione ha obbligato a pensare ad attrezzature meno costose, favorendo scelte che risultano, non solo più economiche, ma soprattutto capaci di rispondere alle sfide di un materiale che mal si presta ad operazioni di scansione massiva. Infine, la mancanza di personale specializzato ha aperto le porte dell'Archivio a tirocinanti provenienti principalmente dell'Università degli studi di Roma "La Sapienza". Grazie a ciò, l'APUG, inserito in una struttura universitaria, non è più solo un centro di documentazione ma anche un centro di formazione.



Nell'ambito degli archivi, in particolare di quelli ecclesiastici, si è passati da un sistema di "segretismo", in parte provocato dal confronto ideologico che dalla metà del XIX secolo si estese fino alla caduta del muro di Berlino, a un sistema sempre più aperto. Se prima la documentazione venne, in molti casi, celata, oggi si registra una disponibilità di documentazione che non sempre contribuisce ad accrescere le conoscenze. Il tipo di *segretismo* che si perpetua ancora oggi non è più legato alla volontà dell'archivista o dell'istituzione ma provocato da strumenti di ricerca inadatti che possono deviare lo studio, da un'assenza di dialogo tra archivisti e ricercatori e dalla carenza di politiche di conservazione che spesso impediscono la consultazione.

Figura 1. Sforza Pallavicino, *Metaphisica*, metà XVII sec. (APUG 390). Illustrazione del celebre calligrafo Galienus Conocius

<sup>4</sup> Per informazioni dettagliate su PK e PKT: <http://Pinakes.imss.fi.it>; <http://Pinakes.imss.fi.it/p3wiki>.

Per valorizzare queste testimonianze documentarie e consentire alla comunità scientifica di accedere a un patrimonio ancora in gran parte inedito l'APUG ha avviato diversi progetti: dalla basilare realizzazione di un OPAC a quella di un Archivio digitale. Aderendo al progetto di censimento nazionale MANUS online<sup>5</sup> dell'ICCU è iniziato nel 2008 il processo di catalogazione dei fondi storici dell'APUG. Questa costituisce la prima esperienza di realizzazione di un OPAC, relativo al materiale documentario manoscritto, all'interno della Compagnia di Gesù. I fondi attualmente in fase di inserimento a catalogo sono: il Fondo Curia, contenente circa 2.380 volumi (sec. XVI-XIX), restituito alla Compagnia di Gesù dalla Bncr nel 1948; il Fondo APUG, costituito da 3.271 volumi (sec. XVI-XIX), in parte miscelanei e provenienti dal Collegio Romano e da precedenti sedi dell'Università fino al 1946; il Fondo Collegio Romano, con testimonianze delle vicende istituzionali dell'Università nel XIX secolo. Grazie alla possibilità data dal *software* di gestire diversi livelli di descrizione (scheda breve/scheda estesa) si sta procedendo in due direzioni: da una parte vengono inseriti i dati ricavati da un'analisi codicologica e filologica approfondita, dall'altra vengono recuperati quelli derivanti dalle schedine dei cataloghi cartacei esistenti. Il progetto dell'ICCU, integrato a MANUS, di realizzazione di un Authority file all'interno di MANUS ha permesso la creazione, per ciascun tipo di responsabilità individuata sui manoscritti, di Authority record (corredati di notizie biografiche e bibliografiche dettagliate) relativi ai Gesuiti. L'APUG diventa così il principale punto di riferimento in materia gesuitica per gli archivi e le biblioteche partecipanti al progetto nazionale. Un catalogo unico permette, infatti, il ricongiungimento virtuale tra i documenti dell'archivio e quelli di altre sedi che conservano fondi gesuitici e la possibilità di confronto tra archivisti e bibliotecari di istituzioni diverse. Durante la catalogazione, inoltre, viene rilevato lo stato di conservazione dei manoscritti che, come si vedrà, è l'azione che deve precedere ogni intervento di digitalizzazione.

Accanto a quest'attività che consente all'archivio di essere presente in un catalogo unico con alcune tra le più importanti istituzioni culturali italiane (aderiscono attualmente al progetto oltre 150 enti), l'APUG ha avviato un progetto per la costituzione di un archivio digitale che valorizzi la complessità del suo patrimonio.

La parte più significativa del materiale documentario conservato presso l'archivio non rientra nelle normali casistiche cui si è abituati nell'ambito della produzione manoscritta. Questa documentazione si sta rivelando preziosa per una ricostruzione della storia dell'insegnamento nell'università che elaborò una consistente razionalizzazione degli studi<sup>6</sup>.

<sup>5</sup> Roberto Marcuccio, *Catalogare e fare ricerca con Manus Online*, «Biblioteche oggi», 28 (2010), n. 6, p. 33-49. Per accedere ai Fondi dell'APUG in MANUS: [http://manus.iccu.sbn.it//opac\\_SchedaBiblioteca.php?ID=168](http://manus.iccu.sbn.it//opac_SchedaBiblioteca.php?ID=168).

<sup>6</sup> Gesuiti, *Ratio atque institutio studiorum Societatis Iesu*, introduzione e traduzione di Angelo Bianchi, Milano: Rizzoli, 2002.

Da documenti<sup>7</sup> conservati in APUG emerge come alcuni manoscritti o stampati, fossero in possesso dei professori gesuiti sino alla loro morte o tramandati ai successori della cattedra per essere poi depositati nella “Camera dei manoscritti”<sup>8</sup>. All’interno della documentazione prodotta dai professori del Collegio Romano, spesso dettagliata con l’indicazione della data in cui si tennero le lezioni, è possibile individuare interventi successivi della stessa mano o di mani differenti che per decenni lavorarono sullo stesso materiale: in alcuni casi l’elemento di interesse si trova non tanto nel testo ma proprio in quelle piccole note a margine. L’indicizzazione di questi elementi, considerati normalmente secondari, è di grande rilevanza per lo studio non solo dell’attività del singolo gesuita, ma per analizzare la struttura e lo svolgimento dei singoli insegnamenti e, più in generale, le rotture e le continuità all’interno del sistema didattico del Collegio Romano. La storia dell’uso e del riutilizzo di materiali può diventare il fulcro di una ricerca che prenda in considerazione la rete di relazioni intervenute sia all’interno di un singolo manoscritto che in un *corpus* di materiale documentario più vasto. Per offrire uno strumento che superi i limiti della ricerca tradizionale l’APUG ha firmato, con la Fondazione rinascimento digitale di Firenze e l’ILC di Pisa, un accordo di collaborazione che consentirà:

- la digitalizzazione integrale di alcune opere. I documenti saranno scelti in base a interventi di valorizzazione specifici: i due grandi progetti in corso riguardano il materiale documentario del Fondo Clavius e il celebre Epistolario di Athanasius Kircher, già digitalizzato in precedenza;
- la realizzazione di un archivio di conservazione dei materiali digitali;
- la fruizione dei dati attraverso l’impiego dell’applicazione Web open source PK, in cui confluiranno i documenti digitalizzati corredati di un apparato di metadati tradizionali, di trascrizioni del testo nelle sue diverse versioni e di descrittori semantici nelle diverse lingue.

L’intervento di digitalizzazione è volto a limitare l’accesso diretto al documento ai soli ricercatori che svolgono ricerche codicologiche, fornendo agli altri l’accesso al contenuto testuale in formato digitale. La digitalizzazione ha senso, infatti, solo in funzione di una conservazione materiale della documentazione originale. Pertanto sono stati affrontati i rischi, spesso sottovalutati, connessi alla movimentazione del materiale documentario, allo stress provocato dall’attività di riproduzione e alla scelta di apparecchiature non adatte. Per queste ragioni le diverse fasi del processo saranno gestite all’interno del laboratorio dell’APUG con attrezzature (*book*

<sup>7</sup> *Nota delle opere composte dal P. Nicolò M.a Pallavicino trovate dopo la sua morte alcune latine, alcune italiane*, carta inserita infine a F.C. 1088.

<sup>8</sup> *Antiche Consuetudini Della Comp.a; massime del Coll.o Rom.o raccolte da libri stampati o dalla tradizione di Padri che vissero prima della Soppressione etc.*, Fal. R, Archivio PUG Storia, doc. 8.

*cradle* corredato di un'apparecchiatura professionale per la fotografia digitale) realizzate secondo le specifiche esigenze dell'archivio: preservare materiale molto fragile, con legature che in molti casi non consentono l'utilizzo di scanner. Ogni fase del lavoro sarà eseguita da personale debitamente formato che si atterrà al protocollo elaborato per questo progetto. Attraverso un *software* di ritocco delle immagini, potranno essere realizzati interventi di restauro virtuale, producendo un oggetto digitale che consenta la miglior leggibilità di documenti che, spesso, presentano inchiostri sbiaditi o fortemente imbruniti. I materiali digitali prodotti, nel formato di conservazione TIFF, saranno a loro volta preservati in un *server* dedicato.

Per quanto riguarda gli strumenti di ricerca, essendo normalmente pensati da chi è abituato a incasellare staticamente dei dati in una griglia, non sempre corrispondono alle necessità pratiche di chi, talvolta, non ha le competenze necessarie per districarsi nel mondo creato da archivisti, bibliotecari o catalogatori, rischiando di conseguenza di non trovare dati fondamentali per il proprio lavoro.

Questo sistema aperto si propone di andare oltre i convenzionali cataloghi di collezioni digitali ed è per tale ragione che sarà necessario partire dall'assunto che esistono diversi livelli di osservazione.

Il concetto di osservazione, mutuato dalla teoria dei sistemi sociali di Niklas Luhmann, potrebbe fungere da cornice teorica nella quale pensare uno strumento di ricerca adatto a un'euristica che tenga conto delle complessità tanto del materiale che della ricerca. L'osservazione implica l'unità, in una stessa operazione, del distinguere e indicare: conoscere è marcare un lato e non un altro in quello che si osserva. Assumendo quindi come valida l'affermazione secondo la quale la realtà è sempre una realtà osservata, il documento, a sua volta testimonianza di un'osservazione, dovrebbe essere descritto lasciando aperta l'analisi alle nuove possibili differenziazioni.

Nell'individuazione dei casi d'uso non sarà sufficiente indicare gli attori che potenzialmente useranno quest'applicazione ma, nel panorama delle risorse online, è più che mai necessario chiedersi quali ricerche sono necessarie e in che modo possono essere svolte dall'utente remoto.

Attingere ad una conoscenza non significa che l'oggetto dell'analisi ci possa essere svelato nella sua "realtà", bensì che attraverso "l'invenzione" di dati esterni gli vengano attribuiti descrittori che lo rendano in qualche misura conoscibile.

I tempi della documentazione archivistica non sono i nostri tempi: la creazione di uno strumento di ricerca deve tenere conto della conflittualità tra il regime di storicità nel quale il manoscritto è stato prodotto e l'aspettativa di un risultato che si pretende, da quando la ricerca avviene attraverso la rete, praticamente immediato. Lo scotto da pagare per questa immediatezza può essere un esito della ricerca incongruente, insufficiente o troppo generico.

Si deve inoltre considerare come la descrizione di ciò che non è contemporaneo crei soltanto un'illusione di contemporaneità, attribuendo a un documento un valore che poteva originariamente non avere.

In un sistema aperto tutti i dati potranno essere integrati e costantemente aggiornati dagli utenti stessi, una volta abilitati, in modo da creare una piattaforma di lavoro che veda la collaborazione della comunità scientifica internazionale. Per garantire un uso appropriato dei dati relativi alla documentazione l'accesso alla banca dati online avverrà in modalità controllata definendo le attività dello specifico utente, monitorandone le ricerche in modo da evitare duplicazioni inutili e ridondanti di edizioni relative ai medesimi manoscritti.

Il *medium* del computer permette di differenziare, nella comunicazione, l'immissione di dati dalla richiesta di informazioni. Come nel caso della scrittura, non c'è unità tra immissione dei dati e comprensione. All'ampliamento della possibilità di gestire un numero sempre crescente di dati e relazioni dovrà accompagnarsi la necessità di una selezione che influirà sugli stessi contenuti.

PK consentirà la gestione di materiale eterogeneo a diversi livelli di descrizione: si pensi al caso di un testo a stampa glossato dall'autore in previsione di una ristampa dove compaiono anche note dei censori, glosse successive e si trovino inserite lettere relative proprio a quell'edizione. In questo caso i livelli di descrizione sono

almeno quattro: il testo edito, le glosse di commento, l'intervento dei censori e il carteggio.

Ovviamente questi diversi livelli prevedono l'utilizzo di standard diversi come, ad esempio, gli *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)*, ISBD(A), o la *Guida a una descrizione uniforme dei manoscritti e al loro censimento* che non entreranno in conflitto, ma dovranno integrarsi a seconda che la descrizione riguardi, nel caso specifico, materiale a stampa o manoscritto.

Attualmente nel caso di te-

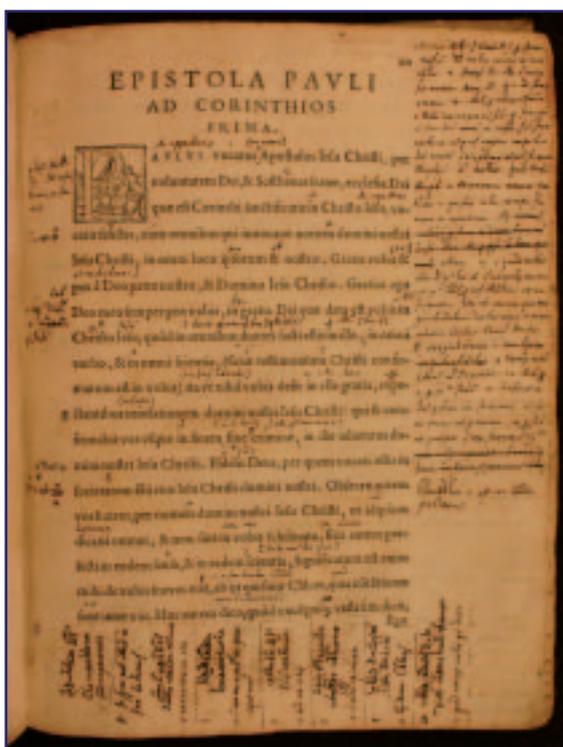


Figura 2. *Biblia. N.T. Ad Corinthios Epistula*, metà XVI sec. (APUG 429). Cinquecentina attribuibile al tipografo Bartholomaeus Gravius riccamente glossata da più mani e in più lingue

sti ibridi, stampa-manoscritto, si può procedere solo in due direzioni: o si descrive l'edizione inserendo come note di esemplare i riferimenti alla parte manoscritta, o si analizza il manoscritto indicando l'edizione come parti a stampa dove dati quali il titolo e i riferimenti alla pubblicazione non sono indicizzati. È evidente invece come in questo oggetto documentario non possano essere scisse le due descrizioni: soltanto analizzando i diversi livelli (stampa/manoscritto) e le relazioni intervenute (glosse/censura) su questo manufatto sarà possibile restituirne la complessità, parzialmente percepita attraverso l'immagine digitale.

Un altro esempio è quello di una legatura di riutilizzo costituita dal foglio di un manoscritto medievale: al singolo oggetto dovranno corrispondere sia i descrittori relativi alla legatura che quelli riguardanti la parte del manoscritto che, se identificata e collegata ad altre legature o a manoscritti mutili, consentirà di ricomporre virtualmente un documento oggi smembrato.

Come sottolineato all'inizio l'immissione di dati che descrivono una tale complessità, se non corredata da strumenti avanzati per la ricerca, potrebbe rivelarsi inutile. È per questo che gran parte del lavoro riguarderà l'elaborazione di una metodologia che consenta nuovi tipi di ricerca nel panorama del Web semantico. L'analisi di questi processi sarà approfondita nei due successivi interventi.

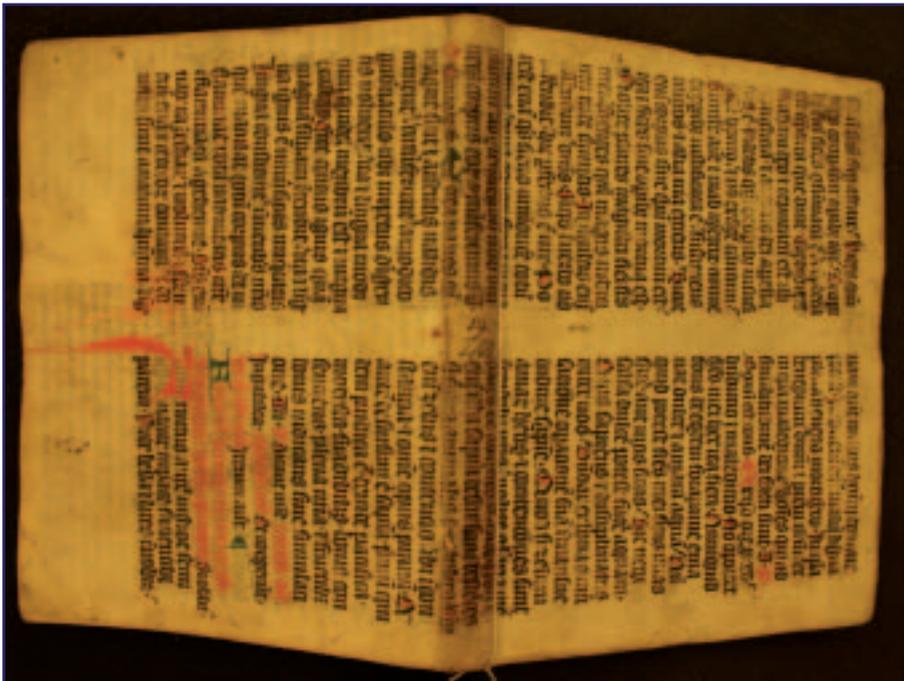


Figura 3. Pergamena medievale riutilizzata come legatura del codice F.C. 462

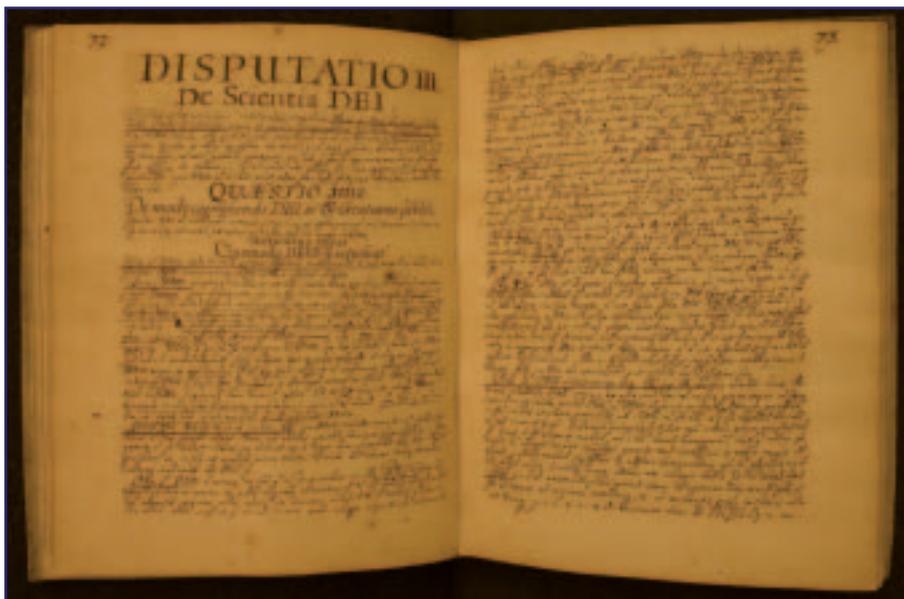


Figura 4. Martinus Juranich, *Tractatus de Deo uno et trino*, 1659 (F.C. 462, p. 72-73)

## Il metodo<sup>9</sup>

In questa sezione prenderò in esame gli aspetti di carattere metodologico che devono essere considerati nel momento in cui un'istituzione culturale (biblioteca, archivio, museo ecc.) diventa digitale e si pone l'obiettivo di rendere fruibili le fonti, in particolare quelle manoscritte inedite, ad un pubblico rappresentato principalmente da specialisti. Vale la pena di evidenziare come il processo di diffusione della cultura digitale (iniziato ormai da molti anni anche se proseguito, specialmente in Italia, con una maggiore lentezza e mancanza di coordinamento rispetto ad altri paesi europei, come, per esempio, il Regno Unito<sup>10</sup>) abbia conosciuto proprio di recente una forte accelerazione. Ciò è dovuto, tra l'altro, al fenomeno della diffusione e vendita di libri elettronici che si accompagna alla disponibilità sul mercato dei cosiddetti *ebook reader*. È molto recente la notizia che vede Amazon, leader indiscusso nella vendita di libri mediante ordinazioni online, aver distribuito per la prima volta un numero di testi elettronici superiore a quelli cartacei. La distribuzione di I-Pad da parte di Apple ha intensificato il fenomeno che ora rappresenta un momento di svolta anche per l'editoria italiana<sup>11</sup>.

<sup>9</sup> Andrea Bozzi (Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa).

<sup>10</sup> Si ricorda a tale proposito il pionieristico *Follett Report* che considerava la nuova tecnologia digitale per le biblioteche come strumento essenziale ai fini dell'educazione superiore. Il testo è leggibile all'indirizzo: <http://www.ukoln.ac.uk/services/papers/follett/report/intro.html>.

<sup>11</sup> Il catalogo dei libri elettronici in lingua italiana è ancora piuttosto limitato, ma le grandi case editrici, come, per esempio, il gruppo Mondadori e Feltrinelli, stanno investendo in questo senso.

Non possiamo certo sapere se e in quale misura tale rivoluzione nella diffusione del sapere (peraltro, non confinata al solo settore librario, ma relativa anche a quello musicale e filmico) condiziona anche la produzione di edizioni critiche, ovvero, il mondo della filologia e dell'editoria scientifica. Nel momento attuale, trovandoci di fronte ad un archivio di fonti con un valore culturale molto significativo e sottoposto a conversione digitale anche per aspetti di conservazione ibrida<sup>12</sup> e di modalità alternative di fruizione, dobbiamo stabilire quali modalità siano le più appropriate affinché l'Archivio storico della Pontificia Università Gregoriana (APUG)<sup>13</sup> possa mettere in valore il posseduto in fase di digitalizzazione indipendentemente dal fatto che tale documentazione sia poi eventualmente distribuita, secondo una non ancora precisabile modalità, su apparecchi portatili come i sopra menzionati *ebook reader* o su macchine collegate ad Internet e dotate di funzioni di annotazione e di ricerca sofisticate.

Partendo da queste premesse, dunque, conviene distinguere due momenti: un primo momento corrisponde a quello che deve collegare il progetto dell'APUG, all'ambiente delle *digital humanities* ed è quello sul quale mi soffermerò; un secondo momento, invece, è quello che mette in relazione i prodotti ottenuti nell'ambito di iniziative come questa con le nuove modalità di diffusione massiva dei dati digitali e che, almeno al momento attuale, non conviene affrontare. Il primo vede principalmente un utilizzo in ambito specialistico, mentre il secondo presuppone la presenza di un pubblico di lettori/utilizzatori molto vasto che viene invogliato ad accedere ai documenti anche per la presenza di elementi di assistenza aggiuntivi alla lettura ed alla comprensione (annotazioni, illustrazioni, note esplicative, accompagnamento sonoro, brevi filmati, ecc.). Potrei concludere questa parte dicendo che la prima fase riguarda la ricerca scientifica e la componente critico-editoriale della documentazione digitale, mentre la seconda riguarderebbe, gli aspetti didattici (strumenti di lettura e ricerca documentaria) che, ove opportunamente controllati, potrebbero valorizzare i percorsi formativi degli allievi delle scuole, specialmente a livello universitario.

Vediamo con maggiori dettagli gli aspetti di carattere metodologico sottesi al sistema informatico per la produzione della biblioteca digitale da parte dell'APUG

<sup>12</sup> La conservazione ibrida contempla diverse modalità di riproduzione di uno stesso documento in modo tale che di esso esistano più copie ottenute con tecnologie diverse (fotografiche e digitali), con lo scopo ambizioso di favorire la creazione di un sistema informativo integrato nel quale tradizione ed innovazione possano cooperare.

<sup>13</sup> Un accordo di collaborazione scientifica e tecnologica è stato firmato fra l'APUG, l'Istituto di linguistica computazionale "A. Zampolli" del CNR di Pisa (ILC) e la Fondazione rinascimento digitale di Firenze (FRD) proprio al fine di sviluppare metodi e strumenti adeguati alla valorizzazione del patrimonio librario storico posseduto dall'istituzione universitaria romana. Date le caratteristiche, descritte nella prima parte di questo contributo, di questa documentazione culturale si è resa necessaria un'attenta valutazione di quanto già realizzato a Pisa e Firenze affinché il sistema sia calibrato sulle specifiche esigenze di catalogazione, conservazione e fruizione dell'APUG.

con caratteristiche tali da garantirne lo studio, l'annotazione e, in una prospettiva ancora più ampia, la condivisione nell'ambito di un'infrastruttura di ricerca per le discipline storiche e filologiche che si ipotizza possa prendere l'avvio nei prossimi anni sulla base di accordi e progetti internazionali sostenuti dalla Commissione Europea e controllati dalla Fondazione Europea della Scienza di Strasburgo. Questo aspetto non va sottovalutato poiché appare sempre più necessaria una visione coordinata di tutte le attività che ruotano intorno alle *digital humanities* affinché si renda possibile o si semplifichi il processo di interoperabilità fra dati del *cultural heritage*, oggi difficilmente integrabili poiché solo in minima parte immessi sul Web con strumenti e metodi non dichiarati o al di fuori di standard condivisi.

Potremo sintetizzare questo approccio elencando gli elementi principali che sono stati presi in considerazione nella progettazione del sistema PKT, per una descrizione più dettagliata della quale si rinvia alla sezione successiva "Strumenti e funzionalità", utilizzato per la catalogazione, gestione ed interrogazione dei documenti digitalizzati a cura dell'APUG.

Un sistema che consenta di effettuare studi di carattere filologico e critico testuale su documenti digitali necessita di almeno tre componenti: la gestione delle immagini, dei testi e degli apparati (bibliografia, note critiche, annotazioni).

### Le immagini

Oltre ai consueti programmi che consentono di applicare varie tipologie di filtri per agevolare la lettura specialmente di quelle zone interessate da danni e da fenomeni di evanescenza degli inchiostri o acidificazione del supporto cartaceo, è importante disporre anche di strumenti in grado di trovare automaticamente o inquadrare manualmente le "zone parola" all'interno dell'immagine digitale. Questa operazione, non troppo complessa per i libri a stampa antichi con una netta spaziatura fra le parole, risulta utile soprattutto nel caso in cui il testo non debba essere interamente trascritto, ma solo alcuni termini significativi meritino di essere annotati ed indicizzati. Si pensi, per esempio, al caso di documenti nei quali sono presenti illustrazioni, grafici, figure geometriche, formule matematiche o chimiche, miniature che rappresentino il vero oggetto di analisi da parte di uno studioso. Il criterio seguito vuole consentire non solo le operazioni di selezione come quelle sopra citate, ma anche di tracciabilità di tutti coloro che siano in qualche modo intervenuti nelle operazioni di selezione. Viene in tal modo resa concreta una prospettiva di lavoro filologico collaborativo in rete. Abbiamo potuto constatare che una simile funzionalità risulta particolarmente utile nella digitalizzazione di libri di storia della scienza ove sono presenti numerose figure, spesso vero obiettivo di una ricerca e talvolta più importanti del testo che le accompagna poiché su esse si concentra l'attenzione del moderno commentatore ed editore.

## I testi

L'esperienza maturata in numerosi centri specializzati nello sviluppo di applicazioni di *text processing* prosegue ormai da parecchi decenni da quando, cioè, sono state realizzate banche dati testuali col fine di documentare stadi di sviluppo di una lingua o di produrre indagini lessicografiche specifiche. Il metodo da noi seguito per questa componente di PKT non si discosta da questi precedenti, ormai ben documentati, ma intende utilizzare in maniera particolarmente sviluppata il principio secondo il quale un testo sia effettivamente rappresentato da più livelli che devono essere marcati affinché non se ne disperda il contenuto informativo. I fenomeni di intertestualità come, per esempio, le citazioni che un autore riferisce altre sue opere o da quelle di autori diversi, appartengono ad un livello di testo specifico che si differenzia dal livello al quale appartiene il testo dell'autore che cita. Un elemento di granularità ancora maggiore potrebbe rendere necessaria la distinzione fra le citazioni da opere in prosa e quelle in poesia, oppure fra quelle che appartengono ad un periodo determinato rispetto ad altre. Questa serie di fenomeni, se opportunamente marcati, aumenta il livello informativo del testo e consente alla gestione automatica degli indici di tenerne conto. Come accennato in precedenza, un'applicazione orientata al lavoro filologico deve consentire, mediante un'interfaccia per l'utente finale semplice e capace di rispondere alle sue necessità di ricerca, di attivare tutti gli elementi di *mark-up* necessari. Essi, a loro volta, dovranno corrispondere direttamente (seguendo delle precise *guidelines*) o indirettamente (attivando un programma di conversione automatico) a sistemi di codifica del testo entrati nell'uso comune attraverso standard riconosciuti a livello internazionale. Sono proprio questi elementi che danno una garanzia che quanto viene svolto anche da un singolo studioso possa entrare a far parte, se lo si riterrà opportuno, di una comunità di studiosi che condividono programmi e risorse all'interno di una stessa infrastruttura di ricerca<sup>14</sup>.

Non è il caso di soffermarsi su un ulteriore ed importante aspetto correlato all'analisi del testo e, in particolare, alla produzione di indici: il sistema PKT presenta un'architettura che lo rende permeabile alle funzioni prodotte, per esempio, da un sistema di analisi morfologica automatica e lemmatizzazione. Un caso specifico appartenente a questo settore è dato dal sistema LemLat<sup>15</sup> per la produzione di in-

<sup>14</sup> Il tema delle *Research Infrastructures* (RI) si sta rilevando molto importante anche nel settore delle scienze umane dopo essere stato ampiamente discusso e analizzato per le cosiddette scienze dure. Le RI sono considerate un elemento imprescindibile per la costituzione di una *European Research Area* (ERA) nella quale ogni comunità di studiosi condivide standard, dati, sistemi, risultati al fine di competere, grazie a strumenti adeguati alla nuova società della conoscenza, con Paesi molto avanzati nel campo della ricerca scientifica e delle innovazioni tecnologiche. Per ERA si veda [http://ec.europa.eu/research/era/index\\_en.htm](http://ec.europa.eu/research/era/index_en.htm). Per quanto riguarda, invece, le RI si veda <http://cordis.europa.eu/infrastructures/>.

<sup>15</sup> Il sistema è stato originariamente disegnato e sviluppato per produrre l'indice lessicale dei grammatici latini antichi presso l'ILC di Pisa. Nel corso degli anni esso ha subito notevoli

dici lessicali latini, ma sono in fase di realizzazione simili strumenti per altre lingue, come, per esempio, l'arabo e l'occitano medievale.

### Le annotazioni

Le annotazioni sono costituite da più elementi diversi: annotazioni libere o informazioni di apparato critico. Nel primo caso contengono osservazioni personali dell'editore (ma anche di chiunque intenda intervenire, se autorizzato, a chiosare un passo, un'illustrazione, una singola espressione linguistica) e si presentano nella forma di testo non strutturato. Il metodo che abbiamo seguito nella progettazione del nostro sistema di filologia computazionale, anche in seguito a esplicite richieste di ricercatori appartenenti ad ambiti di studio molto diversi fra loro, prende in considerazione anche la possibilità di stabilire delle specifiche ontologie di dominio affinché sia il catalogo dei documenti che il testo da essi veicolato possano venire classificati secondo classi semantiche precostituite all'interno dell'ontologia. È inutile sottolineare che l'operazione di classificazione comporta un aggravio di tempo da parte dell'operatore che annota, anche se l'interfaccia venga predisposto in modo che le classi siano presenti, per esempio, in un menù a tendina dove la scelta è certamente rapida ed è garantita omogeneità di comportamenti. In ogni caso, il tempo e le energie spese nella fase di classificazione sono ben ripagate dalla maggiore quantità di informazioni che si recuperano nel momento in cui si accede al testo anche mediante le sottoclassi dello schema ontologico. Per fare un esempio, si potranno ritrovare tutti i passi dove è attestata la descrizione di parti anatomiche del corpo umano e il sistema restituirà i contesti che descrivono il cranio, le braccia, il tronco, ecc., risultato impossibile da ottenere grazie alla sola lettura di un indice alfabetico delle forme o dei lemmi. Un altro esempio molto interessante riguarda la ricerca linguistica su un *corpus* di opere che trattano, appunto, di fenomeni linguistici: un'ontologia specifica di questo dominio, che potrebbe anche essere relativa solo ad un periodo o ad un singolo grande linguista, consentirebbe di recuperare tutte le parti descrittive di fenomeni grafico-fonetici, presenti eventualmente in una determinata zona geografica. Le annotazioni che si presentano nella forma di un apparato critico servono principalmente per rappresentare almeno tre situazioni: la registrazione di varianti, nel caso in cui un testo sia trasmesso da più testimoni (nel progetto dell'APUG materiale documentario principalmente manoscritto) diversi. Fa parte di questo stesso ambito la registrazione di varianti presenti nelle diverse edizioni a stampa di una medesima opera, caso, talvolta, complicato dalla presenza di annotazioni manoscritte da parte dell'autore stesso che compaiono generalmente nei margini della pagina.

trasformazioni in concomitanza con gli sviluppi della tecnologia. Per informazioni, si veda la sezione *Documents* sul sito <http://www.ilc.cnr.it/lemlat/>, oppure Andrea Bozzi – Giuseppe Cappelli, *A Project for Latin Lexicography: 2. A Latin Morphological Analyzer*, «Computers and the Humanities», 24 (1990), n. 5-6, p. 421-426.

Un secondo caso è relativo ai manoscritti di autori moderni e contemporanei che sono intervenuti a più riprese nella modifica del proprio testo, molto raramente in quelli di autori antichi ove, piuttosto, si rinvengono chiose e commenti di altra mano. Un terzo fenomeno riguarda i testimoni unici, ovvero i documenti che riferiscono un testo non altrimenti noto: anche in questo caso, tuttavia, l'annotazione potrebbe assumere la veste di apparato critico qualora il filologo ravvisi errori banali o sostanziali e ritenga di intervenire restituendo il testo presunto originale con una lezione personale.

Il principio da noi seguito tiene presenti tutti questi aspetti e tende a risolverli applicando un principio unico.

Tali impostazioni metodologiche, in incontri con i referenti dell'APUG, sono state perfezionate *ad hoc* e su queste basi si è avviata la fase di progettazione di componenti software che avranno il compito di produrre i contenuti digitali sul Web e di mettere l'utenza, individuabile in una comunità di specialisti, in condizioni di navigare facilmente su dati complessi e di eseguire operazioni di ricerca intuitive e guidate.

## Strumenti e funzionalità<sup>16</sup>

### Gli strumenti

#### *Il progetto Pinakes*

Il progetto PK è il frutto dell'attività svolta all'interno del dipartimento di ricerca dell'Istituto e museo della storia della scienza di Firenze (ora Museo Galileo<sup>17</sup>). Dal 2006 PK è uno fra i principali progetti di ricerca della Fondazione rinascimento digitale<sup>18</sup>. Nell'anno successivo l'ILC si è associato al lavoro di ricerca, soprattutto per quel che riguarda la gestione del testo digitale.

PK è un'applicazione Web che consente la costruzione di una base di conoscenza distribuita e condivisa basata su ontologie<sup>19</sup> e dispone di uno specifico modulo, PKT per la realizzazione di edizioni digitali di testi critici. PK gestisce dati persistenti avvalendosi di strumenti quali Internet e le ontologie. L'applicazione è aperta, modulare, personalizzabile e gratuitamente scaricabile sotto il controllo della General Public License<sup>20</sup>. Si tratta di un'applicazione in grado di gestire molteplici

<sup>16</sup> Marco Rufino (Fondazione rinascimento digitale).

<sup>17</sup> Istituto e museo di storia della scienza, Firenze, <http://www.imss.fi.it>; Museo Galileo, Firenze, <http://www.museogalileo.it> (1 ottobre 2010).

<sup>18</sup> Fondazione rinascimento digitale, Firenze, <http://www.rinascimento-digitale.it> (1 ottobre 2010).

<sup>19</sup> Cfr. la voce "Ontologia (informatica)" in Wikipedia, l'enciclopedia libera: [http://it.wikipedia.org/wiki/Ontologia\\_\(informatica\)](http://it.wikipedia.org/wiki/Ontologia_(informatica)) (20 ottobre 2010).

Per altre informazioni sulla definizione di ontologia: Tom R. Gruber: *What is an Ontology?*, 1992, <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>, Stanford University: Knowledge System Laboratory, <http://www-ksl.stanford.edu> (1 ottobre 2010).

<sup>20</sup> General Public License, <http://www.gnu.org/licenses/gpl.html> (1 ottobre 2010).

tipologie di dati e oggetti o risorse digitali quali testi non strutturati e strutturati (HTML, XML, etc.), immagini, audio, audiovideo, oggetti tridimensionali, tabelle e grafici, espressioni matematiche, GIS, mappe e cartografie. Inoltre, essa consente una gestione multilingue: al momento attuale è possibile utilizzare l'italiano, l'inglese, il francese e il tedesco.

La versione Pinakes 3.0 è formata da tre componenti principali:

- Pinakes Admin: interfaccia Web di amministrazione per la definizione dei progetti, dello schema dei dati, degli utenti e per la configurazione di uno specifico *repository*;
- Pinakes Input: applicazione Web per il popolamento del *repository* e l'*editing* controllato dei dati in un modello PK;
- Pinakes Text: applicazione Web per l'edizione critica digitale di documenti testuali.

L'architettura software consiste di moduli Open Source per il cui sviluppo sono stati utilizzati:

- Linguaggio di sviluppo: JAVA (Jdk 1.5 o succ.);
- Servlet Engine: Tomcat 5.5.x + APACHE HTTP connectors;
- Web Server: APACHE HTTPD server 2.2.x;
- WebApplication Framework: Jakarta Struts; JBoss Seam;
- Web Service Framework: APACHE Axis 1.4;
- Database Engine: PostgreSQL 8.1 o successive;
- IDE: Netbeans 5.5.1.

### *Perché Pinakes*

PK nasce per dare una risposta alle problematiche e metodologie emerse dall'utilizzo dell'IT (Information Technology) per le ricerche e gli studi in ambito umanistico. Infatti, se da un lato l'introduzione dell'IT permette lo sviluppo di metodi formali all'interno delle discipline umanistiche, dall'altro richiede una significativa evoluzione delle metodologie di ricerca, fatto che tarda a realizzarsi anche a causa delle numerose ed ingiustificate resistenze. Tale ritardo influenza negativamente la possibilità di realizzare, comunicare e condividere il sapere attraverso le tecnologie del Web, cioè di individuare modalità ottimali di impiego delle innovazioni, sviluppando uno specifico *data modeling*.

Infatti, mentre sempre più spesso i risultati dell'attività di studio in ambito umanistico vengono codificati in formato digitale, l'utilizzo di differenti modelli per l'organizzazione di questi dati rende difficile, anche nell'ambito della stessa disciplina, lo scambio, l'interoperabilità, la condivisione ed il riuso dei risultati prodotti dall'attività della ricerca.

La caratteristica fondamentale di PK è quella di consentire la costruzione di una base di conoscenza distribuita e condivisa basata su ontologie. A questo PK affianca il fatto di essere un'applicazione Web, e quindi di permettere l'accesso per la modifica, l'implementazione e la consultazione dei dati senza limitazioni di tempo e di luogo.

Queste due caratteristiche garantiscono notevoli vantaggi:

- migliorano l'interscambio dei dati fra le persone e all'interno di istituzioni, enti e organizzazioni perché favorisce l'interoperabilità fra sistemi;
- permettono la creazione di un *data modeling* interoperabile per una gestione integrata ed evoluta dei dati;
- favoriscono, grazie ad una base di conoscenza comune, la condivisione e la ri-usabilità di metodi di modellizzazione, paradigmi, linguaggi e dati;
- permettono la pubblicazione sul Web dei dati provenienti da progetti differenti o da settori disciplinari diversi;
- facilitano il legame di dati contenuti in *repository* o basi dati diverse nonché di tipologie diverse di oggetti o risorse digitali;
- migliorano la ricercabilità dei dati poiché consentono interrogazioni semantiche;
- consentono il riutilizzo di dati digitali già presenti sia sul Web che all'interno di basi dati;
- consentono di riconoscere Proprietà Intellettuale e copyright dei risultati individuali;
- semplificano le collaborazioni interdisciplinari grazie ad un ambiente multilingue;
- offrono una nuova prospettiva e visione sulle relazioni fra i concetti, elemento indispensabile per aprire nuove strade ad ulteriori studi e, quindi, acquisire nuove conoscenze.

### *Ontologie e Web semantico*

Il termine Web semantico<sup>21</sup> indica un'evoluzione del Web attuale in cui le informazioni e i dati assumono una precisa caratterizzazione semantica che rappresenta la sostanziale novità rispetto al Web attuale. Esso renderà possibili elaborazioni au-

<sup>21</sup> Il World Wide Web, come si presenta oggi, non dispone di strumenti in grado di fornire informazioni *ad hoc*: ricerche effettuate sui motori di ricerca, Google *in primis*, solitamente forniscono una grande quantità di risultati, moltissimi dei quali con scarsa attinenza alla richiesta effettuata. Non esiste infatti uno strumento informatico capace di "comprendere" il contenuto di una pagina Internet: il Web, cioè, opera ancora in modo "letterale" (sulla base di ricerca di specifiche stringhe) e non dispone ancora di strumenti capaci di comprendere il valore "semantico" delle interrogazioni.

Questa limitazione è comune sia alla rete di Internet che alle Intranet di enti, istituzioni ed aziende, i cui documenti, dati e informazioni, che rappresentano la "conoscenza" specifica di ciascuna organizzazione, sono sì memorizzati in formato elettronico, ma in modo tale da restare quasi altrettanto inaccessibili di quando venivano archiviati su carta. Infatti, gran parte del contenuto di Internet è ancora progettato per essere "letto" piuttosto che per essere "elaborato" da programmi.

tomatiche molto più complesse ed efficaci. Per tale ragione assume un ruolo determinante la produzione di risorse semanticamente marcate e, soprattutto, l'attribuzione di una connotazione semantica a quelle già esistenti.

PK nasce proprio con questo intento: offrire gli strumenti per la costruzione di una base di conoscenza distribuita e condivisa basata su dati semantici e ontologie; esso dispone di meccanismi grazie ai quali i dati semantici possono essere contenuti da un modello ontologico esterno alla risorsa di riferimento.

La scelta di basare PK su ontologie è derivata dalla considerazione che l'uso di ontologie e le sottostanti tecnologie di modellazione contribuiscono all'organizzazione di architetture software con elevate capacità di interoperabilità, cooperazione, adattamento e capacità di evolvere.

In questa prospettiva, l'approccio ontologico comporta numerose opportunità:

- le ontologie permettono la rappresentazione esplicita di modelli semantici che combinano la non ambiguità necessaria alla specificità tecnica con la comprensibilità capace di colmare la distanza tra gli specialisti e i tecnici;
- il modello ontologico si adatta bene ad un contesto distribuito e permette la creazione di modelli per riuso, composizione e riconciliazione di frammenti sviluppati in modo concorrente e distribuito;
- le ontologie hanno la capacità di modellare domini che evolvono nel tempo, superando molte complessità che invece discendono dall'uso esclusivo di tecnologie di rappresentazione convenzionale. Questo, a sua volta, favorisce un approccio incrementale allo sviluppo capace di accompagnare processi di evoluzione verso modelli condivisi.

È quindi necessario un approccio che renda possibile in forma automatizzata un'interpretazione e una comprensione semantica dei dati. Qualcosa di profondamente diverso dalla nuova Babele di informazioni e linguaggi nella quale è facile perdersi: uno strumento, che può essere rappresentato dal Semantic Web (Web semantico) capace di far ottenere, ad ogni ricerca, risposte mirate e selettive. Il termine Semantic Web è stato proposto per la prima volta da Tim Berners-Lee (Cfr. Tim Berners-Lee – James Hendler – Ora Lassila, 2001: *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a devolution of new possibilities*, «Scientific American», may 2001. Tim Berners-Lee ha sottolineato che uno degli elementi fondamentali del web semantico sarà la compresenza di più ontologie. Se si vuole un sistema dinamico in grado di raffinarsi e funzionare su scala universale, bisognerà pagare il prezzo di una certa dose d'incoerenza. Gli obiettivi del Web Semantico sono di riportare chiarezza, formalità e organizzazione nei dati, collegando l'informazione presente nelle pagine Web a concetti astratti organizzati in una gerarchia (ontologia), a sua volta descritta in un meta-documento; permettendo a vari agenti intelligenti, di interpretare le informazioni e di addentrarsi nella rete navigando come esseri umani. Offrire, quindi, la possibilità di cogliere il contesto semantico di una fonte informativa interpretando le varie relazioni esistenti tra le risorse, formulare asserzioni sulle stesse, nonché controllare la loro attendibilità.

Per approfondire l'argomento: *W3C Semantic Web Activity*, <http://www.w3.org/2001/sw/> (1 ottobre 2010).

L'abituale organizzazione dei dati fatta ricorrendo al meccanismo della classificazione, può presentare diversi problemi derivanti essenzialmente dalle differenze di dominio, terminologia, scelta e significato delle varie caratteristiche ritenute importanti e dalle differenze fra le relazioni più significative. Anche sottili differenze fra le diverse terminologie risultano importanti e la definizione di vocabolari standard non risolve il problema.

All'opposto, un'ontologia, almeno nel senso in cui questo termine viene impiegato nel settore informatico, è un oggetto specifico, progettato per esprimere i concetti in maniera non ambigua e con un buon grado di copertura e di precisione rispetto al modello inteso. Una classificazione basata su metadati (ad esempio Dublin Core, Metadati amministrativi gestionali – MAG), pur essendo finalizzata all'integrazione dei dati, risente sempre della soggettività dei compilatori, mentre un'ontologia rappresenta un modello formale capace di accettare un livello di complessità maggiore, privilegiando la completezza e la correttezza logica rispetto alla comprensibilità umana. Tutto ciò offre una serie di evidenti vantaggi, ma richiede un lungo e profondo lavoro di riflessione per passare da una struttura lineare o relazionale (propria ad esempio delle schede per la catalogazione bibliografica, archivistica, artistica, ecc.) ad una struttura ontologica (e quindi gerarchica) per la costruzione di una base di conoscenza distribuita e condivisa.

### *L'Archivio storico della Pontificia Università Gregoriana*

Come ben illustrato nella prima parte di questo articolo, l'APUG costituisce una fonte archivistica di grande valore che rappresenta, per la sua complessità intrinseca, articolazione e varietà del materiale documentale posseduto, un'appassionante sfida per qualsiasi archivista.

Una sfida che si svolge su due piani: il primo, con caratteristiche più "materiali", sarà quello di rappresentare compiutamente la complessità della documentazione conservata presso l'archivio che, per la maggior parte, non rientra nelle normali casistiche della produzione manoscritta; il secondo, ancor più impegnativo, riuscire a "ricostruire", partendo, dalla utilizzazione continuativa della produzione manoscritta e dalle rielaborazioni successive, la storia dell'insegnamento e della sua evoluzione all'interno del Collegio Romano.

Questa prova assume una dimensione particolare per chi si propone, come PK, di rappresentare in modo logico e semanticamente definito uno specifico dominio di conoscenza, fornendo nel contempo strumenti automatizzati di accesso e consultazione, ma anche di analisi, annotazione e ricerca scientifica collaborativa.

Riteniamo che il progetto dell'APUG per la costituzione di un archivio digitale possa rappresentare uno dei banchi di prova più significativi per le ambizioni di PK nel proporsi come uno dei migliori e più efficienti strumenti di catalogazione archivistica e di rappresentazione semanticamente connotata della realtà contenuta nel materiale documentale.

### Le funzionalità

I tre componenti principali di PK forniscono all'utente l'insieme di funzionalità necessarie per:

- la definizione via Web dello Schema, dei progetti, degli utenti e la configurazione di un *repository* (Pinakes Admin);
- il popolamento via Web del *repository* e l'*editing* controllato dei dati in un modello PK (Pinakes Input);
- l'edizione critica digitale via Web di documenti testuali (Pinakes Text).

Pinakes Admin rappresenta l'area gestionale dove, sulla base di metodi già disponibili all'interno dell'applicazione (ontologia fondazionale<sup>22</sup>), l'utente può definire:

- quanto sia dettagliata la descrizione degli oggetti che vengono classificati;
- che tipo di relazione deve avere il suo oggetto con gli altri oggetti;
- cosa può o meno venire pubblicato sul Web;
- quali dati possono essere condivisi e quali no;
- chi fa che cosa e con quale livello di autorizzazione.

Pinakes Input rappresenta l'area della classificazione dove l'utente, nel rispetto dei vincoli di Proprietà intellettuale e copyright, può:

- inserire i dati utilizzando una propria definizione e, nel caso di modifiche che si rendano necessarie, come sia possibile effettuarle in base ai metodi adottati;
- raffinare la definizione del dominio di informazione e di conoscenza in base alle necessità incontrate nel corso della ricerca;
- consultare ed utilizzare i dati appartenenti ad altri progetti che potrebbero essere pertinenti alla propria ricerca;
- usufruire e/o offrire altre relazioni di progetto e indici di definizione;
- accedere ai depositi digitali di diversi progetti e utilizzare tutti gli oggetti digitali (immagini, video, musica, ecc.);
- definire una navigazione sui dati con lo scopo di modificarli;
- accedere tramite Web ad altri database allo scopo di catturare dati ed importarli in un progetto esistente;

<sup>22</sup> Con *Ontologia fondazionale* si vuole indicare il "modello concettuale" (*core ontology*) che PK mette a disposizione per aiutare gli utenti nella creazione di ontologie di dominio adatte all'analisi del proprio dominio di conoscenza.

L'ontologia fondazionale di PK fornisce un modello globale estensibile in cui possano essere messi in corrispondenza ed integrati dati provenienti da fonti eterogenee, grazie ad un insieme di classi e di predicati dotati di un significato determinato che possono venire estese dall'utente per poter rappresentare il proprio dominio di conoscenza.

- accedere a depositi digitali esterni al progetto Pinakes 3.0 per reperire nuove risorse o fare riferimento ad esse attraverso la creazione automatica di un indirizzo Web.

### 1. Pinakes Text rappresenta l'area di lavoro sui documenti digitali dove l'utente può:

Ricerca e consultare testi sulla base:

- del titolo;
- della responsabilità (autore, copista, legatore ecc.);
- di eventuali indici specifici (come quello per volumi – nel caso esista);
- del testo: le ricerche all'interno del testo consentiranno di individuare parole (interi o parziali), insiemi semplici o complessi di parole.

In particolare sarà possibile:

- individuare stringhe di caratteri appartenenti ad una parola intera o costituenti la parte iniziale, centrale o finale;
- indicare lo specifico ambito di ricerca, dal *full text* ad uno specifico livello in cui il testo è stato suddiviso (titolo, citazioni, note, etc.);
- attribuire valore distintivo agli accenti, agli spiriti o al maiuscolo;
- cercare una stringa di caratteri in concorrenza o in alternativa ad una seconda stringa, utilizzando operatori booleani di inclusione o esclusione;
- indicare quale intervallo di parole considerare soglia massima nella ricerca di due stringhe di caratteri con operatori booleani.

All'interno della pagina di consultazione sarà disponibile sia l'immagine digitale del documento che la trascrizione del testo in essa leggibile.

### 2. Fornire nuovi contenuti, che potranno aggiungere conoscenza, mediante l'inserimento di nuovi documenti con le relative immagini. A tale scopo si utilizza il linguaggio di *markup* della Text Encoding Initiative (TEI) come standard di riferimento. L'utilizzatore può:

- inserire (o collegare) nuove immagini di pagine di testo;
- inserire (o collegare) le trascrizioni del testo contenuto nelle immagini;
- collegare in forma dinamica il testo della trascrizione all'immagine corrispondente;
- inserire (o collegare) informazioni extra-testuali quali apparati e bibliografia.

### 3. Disporre di funzionalità necessarie alla ricerca filologica, quali:

- inserimento di annotazioni relative al testo;
- consultazione di annotazioni relative al testo;
- confronto e individuazione delle varianti del testo;
- indicizzazione del testo e delle varianti;
- selezione del testo caratterizzato da varianti;
- marcatura delle varianti d'apparato e la visione del testo marcato e della variante;
- lavoro collaborativo realizzato da più utenti in postazioni remote sul medesimo *corpus* di testi e di immagini che sono oggetto di studio, attribuendo a ciascuno studioso la propria specifica responsabilità su quanto elaborato.

Ulteriori funzionalità specialistiche sono inoltre in via di progettazione quali, per esempio:

- la gestione di versioni diverse degli stessi documenti (versionamento);
  - la possibilità di associare un ciclo di vita ai documenti condivisi (nello stesso modo di un articolo scientifico nel quale si riconoscono tre fasi, *draft*, *revision* e *published*) i quali, nelle varie fasi, sono soggetti a vedere modificati i vari livelli di autorizzazione che regolano le modalità e i tipi di intervento.
2. Effettuare annotazioni e, di conseguenza, ricerche semantiche sui testi, con la possibilità di definire un proprio modello concettuale, eventualmente derivandolo da una ontologia fondazionale, e utilizzarlo per descrivere la semantica delle entità presenti in un testo e le relazioni tra testi differenti.

# La digitalizzazione dei Plutei laurenziani: prime considerazioni a conclusione del progetto

**Sabina Magrini**

*Biblioteca statale di Trieste*

**N**el novembre 2010 si sono conclusi i lavori di indicizzazione, digitalizzazione integrale e pubblicazione in Internet di poco più dei 3.900 manoscritti appartenenti al fondo Plutei della Biblioteca Medicea Laurenziana. Si è trattato di un grosso progetto avviato dalla Biblioteca nel giugno del 2008, dopo una lunga fase istruttoria e nell'alveo delle iniziative promosse e finanziate dalla Biblioteca digitale italiana.

Non è facile per me, impegnata a fianco di Emiliano degl'Innocenti (Società internazionale per lo studio del medioevo latino, SISMEL) nella direzione dei lavori del progetto, scriverne ora così a ridosso delle sue ultime battute. Ritengo questo, piuttosto, il momento del silenzio e della riflessione e, soprattutto, dell'ascolto dei pareri, dei suggerimenti e delle critiche dei molti che nel quotidiano delle proprie attività di studio, di ricerca o di divagazione consultano il "prodotto" realizzato dal progetto che è liberamente accessibile in rete all'indirizzo: <http://teca.bmlonline.it>.

Ciò non di meno è pure evidente che una riflessione silente e isolata potrebbe anche, alla lunga, rivelarsi sterile qualora si consideri che proprio le pagine di questa rivista hanno accolto nel 2007, a firma di degl'Innocenti, l'annuncio del progetto e la sua descrizione programmatica<sup>1</sup>. Una valutazione "a caldo" e, ancor più, pubblica dell'eventuale scarto tra ciò che è stato realizzato e quello che si sarebbe dovuto o desiderato mettere in atto mi sembra a questo punto quasi doverosa...

Il capitolato tecnico del progetto elaborato dall'ICCU nel 2005 e approvato dal Centro nazionale per informatica nella pubblica amministrazione (CNIPA) nell'agosto 2006 definiva nettamente caratteristiche e obiettivi dei lavori da eseguire. Al contempo, tuttavia, per la Biblioteca e la SISMEL era altrettanto chiaro che il lavoro in sé, per quanto realizzato "a regola d'arte" da parte dell'associazione temporanea d'impresе (ATI) incaricata<sup>2</sup>, non si sarebbe potuto considerare concluso; esso lo sarebbe risultato solo se fosse stato completamente integrato

<sup>1</sup> Emiliano degl'Innocenti, *Il progetto di digitalizzazione dei Plutei della Biblioteca Medicea Laurenziana di Firenze*, «DigItalia», 1 (2007), p. 103-114.

<sup>2</sup> CESEPI Scarl (Consorzio europeo soluzioni e progetti informatici) – SIAV Spa.

nel contesto delle altre risorse digitali laurenziane – e non solo laurenziane – disponibili in rete. Per la Direzione dei lavori si rendeva, quindi, necessario provvedere non soltanto alla verifica dell'andamento del progetto, ma anche della interoperabilità del suo risultato con il sistema delle risorse informative presenti nel reticolo definito dall'OPAC della Biblioteca (<http://opac.bml.firenze.sbn.it/>)<sup>3</sup> e, grazie al suo corredo di metadati di tipo bibliografico, amministrativo e gestionale<sup>4</sup>, con i maggiori portali culturali nazionali ed europei (ad esempio Internet Culturale, CulturalItalia, Europea).

Da qualche anno a questa parte la Biblioteca Medicea Laurenziana, di fatti, è stata fortemente impegnata nella implementazione, organizzazione, interconnessione e pubblicazione in rete di contenuti digitali (la digitalizzazione e indicizzazione del catalogo Del Furia, la descrizione e la bibliografia dei manoscritti, la scansione delle schede relative alla stessa movimentazione dei codici). È evidente che anche il progetto della digitalizzazione dei Plutei doveva inserirsi a pieno titolo in questo contesto e che anzi, a prescindere dal valore delle sue caratteristiche intrinseche, sarebbe stata proprio la sua perfetta integrazione nell'OPAC laurenziano a costituire uno dei suoi veri punti di forza<sup>5</sup>.

Si intende che nella creazione, raccolta e interconnessione di strumenti e basi dati utili allo studio dei manoscritti quella della Laurenziana non è certo un'esperienza isolata. In questo senso, almeno in Italia, l'impresa del catalogo aperto della Biblioteca Malatestiana di Cesena (<http://www.Malatestiana.it/manoscritti/index.htm>) presentato nel marzo 2003 è stato davvero un apripista<sup>6</sup>. Per la prima volta nel nostro paese, credo, un ente conservatore si è cimentato in prima persona, seppure in collaborazione con forze universitarie, in un progetto di largo respiro per la realizzazione, aggregazione e pubblicazione in rete – tramite il sito istituzionale – di risorse digitali (testi fondamentali per la storia della Malatestiana, descrizioni di manoscritti, una bibliografia continuamente arricchita, immagini) relative al proprio patrimonio manoscritto. Il risultato di questo lavoro, per definizione *in fieri*, consta attualmente in 4.870 voci bibliografiche, 1.399 descrizioni, 42.149 immagini relative ad alcuni dei 429 esemplari manoscritti della Biblioteca.

<sup>3</sup> L'implementazione dell'OPAC è avvenuta in collaborazione con Ifnet Srl.

<sup>4</sup> Lo standard applicato è naturalmente conforme allo standard MAG 2.0, definito dall'ICCU al fine di promuovere l'integrazione e lo scambio delle risorse digitali.

<sup>5</sup> Sull'argomento cfr. anche Sabina Magrini, *Il catalogo aperto della Biblioteca Laurenziana*, «Biblioteche Oggi», 28 (2010), n. 5, p. 11-20.

<sup>6</sup> Antonio Cartelli – Andrea Daltri – Paola Errani – Marco Palma – Paolo Zanfini, *Il catalogo aperto dei manoscritti Malatestiani*, in: *Kodikologie und Paläographie in digitalen Zeitalter*, herausgegeben von Malte Rehbein, Patrick Sahle, Torsten Schaßan, Norderstedt: BoD, 2009, p. 13-23.



Figura 1 a. Catalogo aperto dei manoscritti Malatestiani: Bibliografia



Figura 1 b. Catalogo aperto dei manoscritti Malatestiani: Descrizioni

Anche all'estero sono diversi, anche se non numerosissimi, i progetti con caratteristiche in parte affini. Nella individuazione di quest'ultimi pesa, tuttavia, la oggettiva difficoltà di orientarsi nella miriade di progetti incentrati sullo studio e la valorizzazione dei manoscritti presenti online. Sono tristemente noti i problemi di chi voglia censire imprese di tal genere. Dal momento che le liste dei siti "utili" presenti in rete non sono né esaustive né aggiornate è essenziale basarsi sull'esperienza personale, sul passaparola tra colleghi e sulla verifica costante di notizie o



Figura 1 c. Catalogo aperto dei manoscritti Malatestiani: Immagini

recensioni di vario genere<sup>7</sup>. Nel novero dei progetti a me noti e assimilabili in qualche misura all'impegno laurenziano – ossia di lavori che prevedono, oltre alla digitalizzazione e alla descrizione catalografica di un fondo o di fondi di manoscritti in qualche maniera circoscritti e affini, anche link ad altre risorse interne o esterne – sono da citare, a mio avviso, i casi di Belgica (<http://belgica.kbr.be/>) e di Scriptorium: Medieval and Early Modern Manuscripts Online (<http://scriptorium.english.cam.ac.uk/manuscripts/>) oltre ai ben più famosi Codices Electronici Ecclesiae Coloniensis (<http://www.ceec.uni-koeln.de>)<sup>8</sup> ed E-codices: Virtual Manuscript Library of Switzerland (<http://www.e-codices.unifr.ch>)<sup>9</sup>.

Belgica è la biblioteca digitale implementata dalla Biblioteca reale del Belgio e offre accesso gratuito a diverse categorie di documenti tra i quali una piccola selezione (18) di manoscritti digitalizzati integralmente. Il sito, indirizzato sia ai ricercatori che a un pubblico di amatori e di curiosi, fornisce anche un corredo di notizie (una breve scheda descrittiva del pezzo, una bibliografia, e la possibilità di scaricare in formato .pdf le scansioni dei principali contributi scientifici sull'*item*).

<sup>7</sup> È assai preziosa in quest'ottica – e forse meno nota rispetto ad altre liste affini – la pubblicazione periodica in rete delle attività di digitalizzazione in corso segnalate delle diverse delegazioni nazionali che partecipano all'Association of European Research Libraries (cfr. <http://liber-manuscripts.kb.nl/progressreports/index.html>).

<sup>8</sup> Il progetto ha previsto la digitalizzazione integrale di 424 manoscritti conservati in 9 biblioteche, per un totale di 140.231 pagine, corredati da descrizioni catalografiche e da una sezione separata con bibliografia organizzata per temi e per alcuni manoscritti.

<sup>9</sup> Il progetto ha prodotto la digitalizzazione integrale di 659 codici conservati in 28 biblioteche. Oltre alla descrizione catalografica sono offerti link ad eventuali risorse esterne.

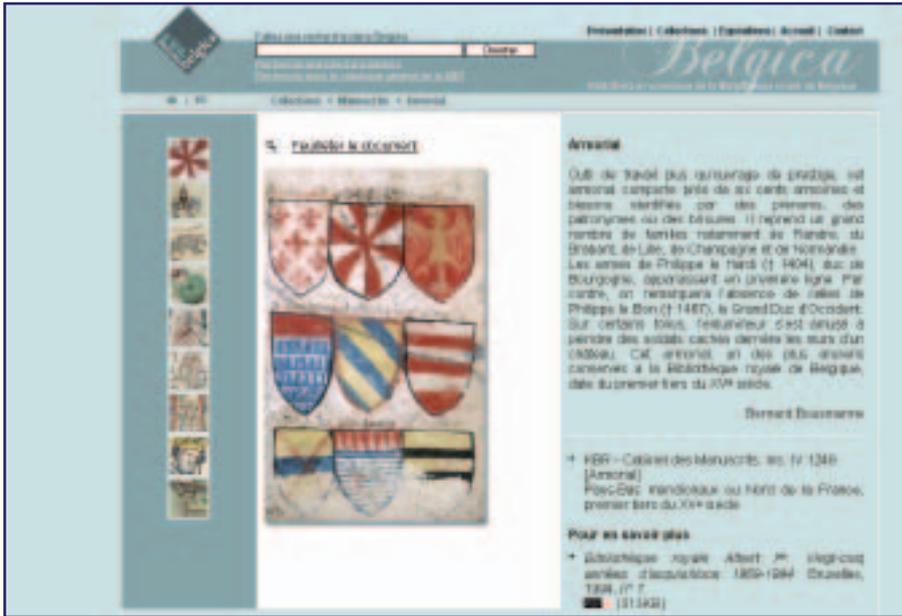


Figura 2. Belgica: Immagini, Materiali vari

Similmente anche il progetto Scriptorium avviato tra il 2006 e il 2009 dall'Università di Cambridge (UK) permette all'utente collegato di visualizzare la digitalizzazione integrale di 185 codici.



Figura 3. Scriptorium: Medieval and Early Modern Manuscripts Online. Immagini, materiali vari

Per ogni manoscritto sono previsti inoltre links a descrizione esterna, interna, bibliografia e alla copia in formato .pdf di un articolo correlato significativo.

È superfluo segnalare che in tutti i casi citati il visualizzatore delle immagini è dotato di strumenti più o meno sofisticati per l'ottimizzazione dei file.

Oltre alla ricerca di una completa integrazione nell'OPAC laurenziano, un'altra caratteristica saliente del progetto di indicizzazione e di digitalizzazione dei Plutei laurenziani è rappresentata quindi anche dalla sua "mole" soprattutto se si confronta con i progetti simili esistenti in rete. A conclusione dei lavori la Teca, accessibile liberamente via internet, ospita un totale di 1.392.984 file immagine in formato .JPEG con risoluzione a 100–150 ppi, di cui 1.386.978 corrispondenti ai manoscritti e 6.006 alle pagine dei cataloghi storici digitalizzati di supporto, per un equivalente di oltre 650 Gigabyte. Tale Teca, a sua volta, va ad integrarsi con l'OPAC della Biblioteca che gestisce ad oggi, tra l'altro, 75.151 abstract bibliografici derivati dall'esame di 4.578 monografie e 13.736 spogli.

È pur vero che l'eventuale successo del progetto laurenziano non dipende affatto dalle sue dimensioni, ma è certo che il numero e la natura particolare dei pezzi da indicizzare e digitalizzare, nonché i tempi stretti in cui si è deciso di operare, hanno fortemente condizionato il *workflow* concordato dalla Biblioteca e dell'ATI vincitrice dell'appalto, nonché imposto una politica attentissima a garantire la stabilità dell'intero sistema.

Le caratteristiche tecniche del progetto sono già state descritte nel dettaglio in corso d'opera<sup>10</sup>; è il caso di tornare sull'argomento, tuttavia, poiché nella fase conclusiva dei lavori è stata introdotta una piccola variante rispetto al capitolato tecnico che può essere di qualche interesse segnalare.

Come è noto, i lavori si sono articolati in tre attività: la acquisizione tramite scanner planetari delle immagini integrali dei manoscritti presso la sede della Biblioteca Medicea Laurenziana, l'indicizzazione con estrazione dei dati bibliografici dai tre cataloghi storici pertinenti il fondo<sup>11</sup> e la pubblicazione in rete delle immagini correlate dai rispettivi metadati MAG (metadati amministrativi gestionali).

<sup>10</sup> Oltre ai contributi già citati alle note n. 1 e 3 cfr. anche Emiliano degl'Innocenti – Sabina Magrini, *Digitizing Cultural Heritage: the Digital Library of the Biblioteca Medicea Laurenziana*, in: *Eva 2009 Florence. Proceedings*, ed. by Vito Cappellini, James Hemsley, Bologna: Pitagora Editrice, 2009, p. 58-63.

<sup>11</sup> Va sottolineato a tale proposito che obiettivo del progetto è sempre stato la sola digitalizzazione dei manoscritti – da realizzarsi entro 30 mesi ed entro un budget di 1.500.000,00 euro – e la sua pubblicazione in rete tramite una teca interrogabile. Ogni ipotesi di catalogazione *ex novo* dei manoscritti è stata pertanto scartata, sia dal Comitato guida della Biblioteca digitale italiana sia dalla Biblioteca Medicea Laurenziana, come improponibile. I cataloghi storici dal quale sono stati ricavati i dati bibliografici utilizzati per la implementazione della teca sono i seguenti: Stefano Evodio Assemani, *Bibliothecae Mediceae Laurentianae et Palatinae codicum mms. Orientalium Catalogus*, Florentiae: Ex typographio Albiziniano, 1742; Anton Maria Biscioni, *Bibliothecae Mediceo-Laurentianae Catalogus* [...], Florentiae: ex Imperiali Typographio, 1752; Angelo Maria Bandini,

Quanto alla digitalizzazione, il capitolato prevedeva inizialmente un'acquisizione di due tipi di file diversi per formato e destinazione (TIFF non compresso a 600 ppi e JPEG compresso a 72 ppi). All'avvio dei lavori la direzione dei lavori ha ottenuto che le apparecchiature di scansione producessero, invece, della medesima immagine tre file (TIFF non compresso a 600 dpi, JPEG compresso a 300 ppi e JPEG compresso a 100 o 150 ppi in considerazione del formato più o meno piccolo dell'*item*). Le prime (le immagini a 600 ppi) costituiscono il formato master, destinato alla conservazione a lungo termine; le seconde (le immagini a 300 ppi), più maneggevoli, sono per la consultazione da parte degli utenti all'interno della intranet della Biblioteca e per la produzione delle copie richieste dagli utenti stessi; le terze (le immagini a 100-150 ppi) per la pubblicazione, accessibile a titolo gratuito, in internet. Nel corso della digitalizzazione dell'ultimo lotto di manoscritti che comprendeva, oltre alle ultime segnature del fondo<sup>12</sup>, l'insieme dei ventotto codici risultati "fuori formato" per le notevoli dimensioni si è resa necessaria un'ulteriore variante nei formati di acquisizione. Si è constatato difatti che di norma i manoscritti "fuori formato" presentavano una altezza superiore ai 450 mm generando così immagini master di "peso" assai difficilmente gestibile in fase di *post processing*. Si è proceduto pertanto a definire due nuove classi di acquisizione: immagini TIFF non compresse a 400 ppi per i codici con altezza compresa tra i mm 450-500 e immagini TIFF non compresse a 300 ppi per manoscritti con altezza superiore ai 500 mm.

L'andamento del progetto ha comportato, quindi, una notevole dose di flessibilità, da parte sia della Biblioteca sia dell'ATI aggiudicataria. Ciò si è reso evidente sin dalle primissime battute dei lavori quando veniva messo a punto il protocollo dell'indicizzazione bibliografica che, per quanto ridotta a "meri" quattro dati (segna-tura, autore, titolo, data) ripresi da cataloghi storici, ha creato da principio qualche intoppo poi brillantemente superato grazie alla disponibilità dell'ATI che ha provveduto a coinvolgere un numero di addetti ben più cospicuo di quelli che aveva previsto inizialmente.

Parimenti l'integrazione nell'OPAC ha previsto un'iniziale fase di assestamento che ha reso necessaria la creazione, da parte dei tecnici variamente coinvolti nel progetto di indicizzazione e digitalizzazione dei Plutei da un lato e nella implementazione del catalogo dall'altro, di una sorta di *cross walk* dei relativi metadati.

Come si è sottolineato, la Teca digitale si integra perfettamente con le altre risorse informative messe a disposizione dall'OPAC della Biblioteca.

*Catalogus codicum manuseriptorum Bibliothecae Mediceae Laurentianae varia continens opera graecorum patrum [...] I-III, Florentiae: Typis Caesareis, 1764-1770; Angelo Maria Bandini, Catalogus codicum latinorum Bibliothecae Mediceae Laurentianae sub auspiciis Petri Leopoldi [...] I-V, Florentiae: s.n., 1774-1778; Angelo Maria Bandini, Bibliotheca Leopoldina Laurentiana, seu, Catalogus manuseriptorum qui iussu Petri Leopoldi [...] I-III, Florentiae: Typis Caesareis, 1791-1793.*

<sup>12</sup> Nella digitalizzazione si è convenuto, infatti, di procedere secondo l'ordine topografico.

L'utente viene messo nelle condizioni di potere passare in modo semplice e veloce dalla visualizzazione del manoscritto alla lettura delle pagine pertinenti del catalogo storico, dall'esame della bibliografia del codice alla disamina di chi lo ha consultato dall'inizio del secolo scorso.

Al contempo gli viene offerta la possibilità di consultare le descrizioni recenti del

**Ricerca Libera** | Ricerca Base | Ricerca Avanzata

Ricerca Libera

perfezionare ricerca

Cerca      Cancella

---

**Risultati Ricerca Libera**

Pagine: 1 2 3 4 5 6 7 8 9 >>      Documenti trovati: **647**      Documenti filtrati per [Datazioni del ms. da catalogo]: 11-20 di

Tutti i Documenti	Unità testuali da catalogo [584]	Intestazioni del ms. da catalogo [63]	
Seignazione	Secolo	Autore	Titolo
PMA.01.01	1481-1500	Agostino	S. Agostino Sermones in evangelio S. Joannis
PMA.01.02	1481-1500	Agostino	S. Agostino De civitate Dei
PMA.01.03	1481-1500	Agostino	S. Agostino Sermones
PMA.01.04	1481-1500	Agostino	S. Agostino Sermones
PMA.01.05	1481-1500	Agostino	S. Agostino Sermones
PMA.01.06	1481-1500	Agostino	S. Agostino De civitate Dei

Figura 4 a. Biblioteca Medicea Laurenziana: Teca – Risultati interrogazione per autore

Figura 4 b. Biblioteca Medicea Laurenziana: Teca – Visualizzazione immagine



Figura 4 c. Biblioteca Medicea Laurenziana: Teca – Visualizzazione catalogo storico



Figura 4 d. Biblioteca Medicea Laurenziana: Teca – Visualizzazione OPAC

manoscritto disponibili o di navigare all'interno di percorsi espositivi virtuali laurenziani (e non) che comprendono quel testimone in particolare. È facile immaginare come un simile strumento possa incontrare il favore del pubblico.



Figura 4 e. Biblioteca Medicea Laurenziana: OPAC – Visualizzazione bibliografia

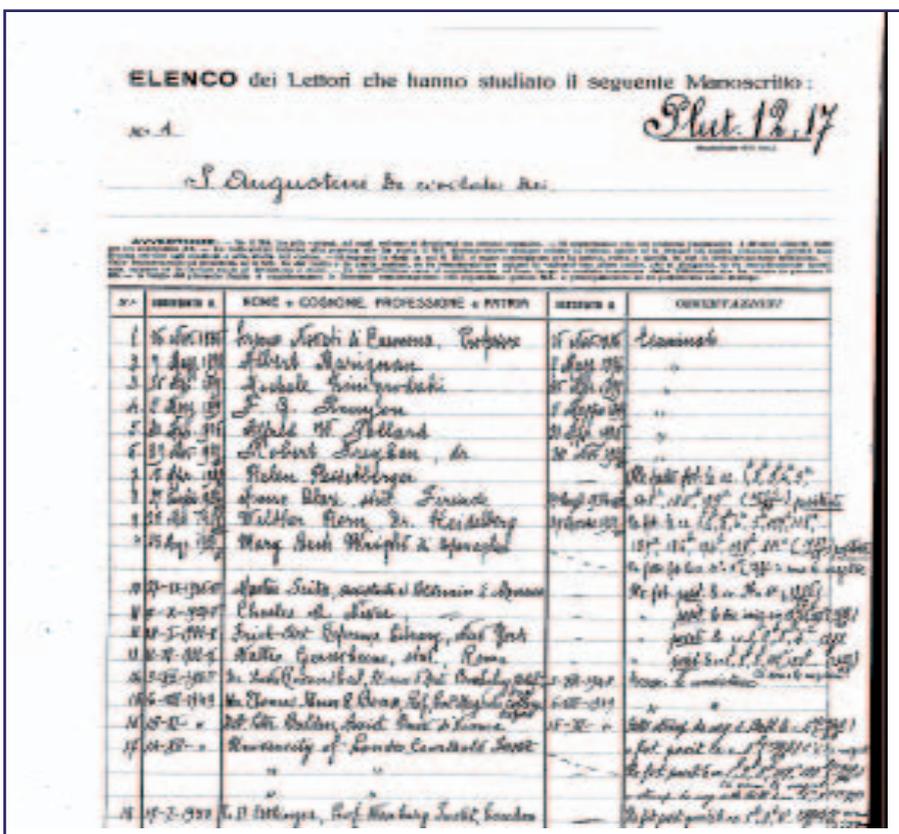


Figura 4 f. Biblioteca Medicea Laurenziana: Teca – Visualizzazione scheda della movimentazione

Già dalla pubblicazione in rete del I lotto del progetto Plutei (606.152 file immagine corredati da relativi metadati e corrispondenti a 1.655 manoscritti e relativi cataloghi storici) che ebbe luogo nel novembre del 2008 il numero dei frequentatori della Teca digitale è stato continuamente in crescita. Si è passati dalle 17.364 richieste con 176 pagine consultate del novembre 2008 alle 172.563 richieste e 4.334 pagine del novembre 2010. Le ricerche vengono effettuate, prevalentemente, per segnatura o autore.

Parallelamente, e in buona parte di riflesso, sono grandemente aumentate le consultazioni dell'OPAC laurenziano: ben 1.293 visite con 15.530 pagine visualizzate nel dicembre 2010. In questo caso il fenomeno è certamente legato all'effetto "trainante" della presenza in rete della pubblicazione delle digitalizzazioni dei Plutei, ma non va sottovalutato anche il grosso impegno profuso nel corso dello stesso 2010 per la creazione di strumenti che rendessero la consultazione di questa piattaforma sempre più immediata e accessibile e che sicuramente ha influito sull'incremento dei contatti. La soddisfazione degli utenti è rilevabile anche tramite i loro interventi diretti: lettere o comunicazioni di apprezzamento. Va notato, tuttavia, che dal novembre 2008 ad oggi almeno una decina di studiosi ha segnalato qualche difficoltà nella visualizzazione delle immagini della Teca, ricollegabili essenzialmente alla necessità da parte dell'utente di installare l'applet Java Runtime Environment (vers. 1.5 o superiore) e alla non corretta gestione della cache di Java. Si tratta di un numero di casi assai esiguo, ma tale da evidenziare una criticità importante del sistema. Per ovviare si è predisposto un *vademecum ad hoc* accessibile dall'homepage ad integrazione dell'*help* già esistente.

Tra gli aspetti più innovativi dell'OPAC laurenziano vi è, tra l'altro, la possibilità per l'utente registrato di pubblicare il proprio contributo (comunicazioni su lavori in corso, lavori originali, testi già pubblicati, notizie, recensioni, commenti informali) in un'area libera, priva di una certificazione calata dall'alto, per così dire, e ben distinta – per non ingenerare confusione circa lo *status* delle notizie – dalla sezione della bibliografia dei manoscritti che è "validata", invece, sia dal processo editoriale subito dagli *items* spogliati sia dall'intervento normalizzatore della Biblioteca stessa.

Sino a qualche mese fa l'area non aveva conosciuto un largo impiego, se non a scopo sperimentale. Nella primavera del 2010, tuttavia, ha preso corpo un'interessante collaborazione tra la Biblioteca Medicea Laurenziana e il Dipartimento di scienze storiche del mondo antico della Facoltà di lettere e filosofia dell'Università degli studi di Pisa. Nell'ambito di un Progetto di ricerca di interesse nazionale (PRIN 2007) cofinanziato dalla stessa Università e dal Ministero dell'istruzione, dell'università e della ricerca, infatti, è stata avviata tra il 2008 e il 2010 la nuova catalogazione dei manoscritti siriaci laurenziani appartenenti al fondo Plutei, ma non soltanto. Si tratta di 70 codici, di cui 50 di notevole rilevanza per le caratteristiche materiali, contenutistiche e storiche: ad esempio il più antico codice illustrato datato (Plut.1.56, Codice di Rabbula) oppure uno tra i più antichi manoscritti biblici completi (Or. 58) traditi, del sec. IX. Unico catalogo a disposizione degli

studiosi per un approccio mediato a questi pezzi è stato sino ad ora quello, ormai assai datato, curato da Stefano Evodio Assemani nel 1742 e semplicemente ripreso da Angelo Maria Bandini nel 1752.

Risultato di questo lavoro sono dunque una nuova descrizione codicologica e contenutistica, una scheda bibliografica laddove possibile, la digitalizzazione integrale per ciascun item<sup>13</sup>. Al fine di ottenere un sensibile risparmio economico e soprattutto una maggiore fruibilità e aggiornabilità, di concerto con la Biblioteca, è stato deciso di pubblicare il catalogo, nelle sue varie componenti, sull'OPAC della Biblioteca: le digitalizzazioni, per un totale di 28.360 immagini<sup>14</sup> e relativi metadati ne popoleranno la Teca; le descrizioni in formato .pdf occuperanno la sezione "Altre fonti descrittive" dell'OPAC mentre le schede bibliografiche, più soggette ad aggiornamento, saranno disponibili nell'area destinata al contributo degli utenti registrati, sempre nell'OPAC.

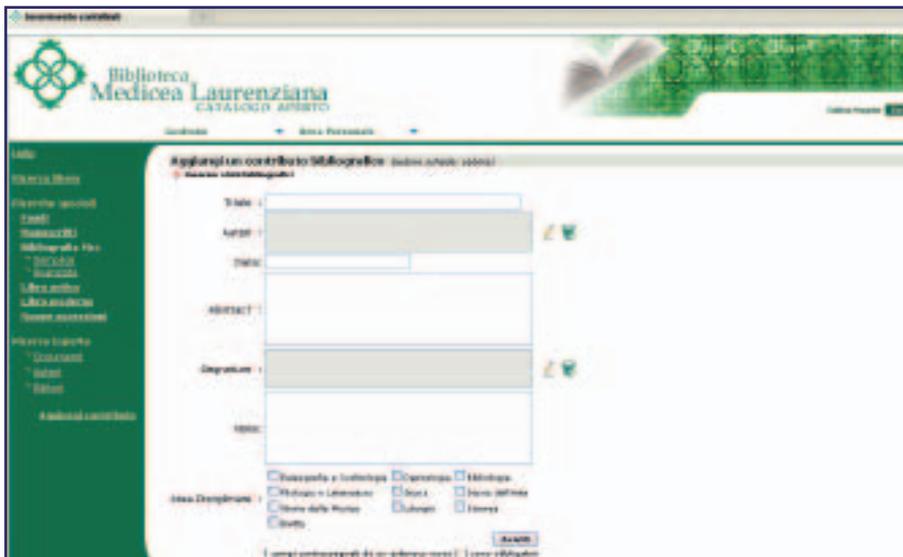


Figura 5. Biblioteca Medicea Laurenziana: Teca – Visualizzazione scheda contributo utente

La portata, a mio avviso, del progetto di digitalizzazione dei Plutei laurenziani appena portato a termine sembra andare ben oltre il risultato programmato in origine. La circostanza non deve meravigliare: un progetto di tal sorta, se davvero riuscito e sostenibile, dovrebbe normalmente catalizzare di fatto una molteplicità di iniziative dalla ricaduta positiva per la Biblioteca e per i suoi utenti.

<sup>13</sup> Il materiale è stato stilato e raccolto da Margherita Farina (Università degli studi di Pisa) e convalidato dal coordinatore del progetto Pier Giorgio Borbone (ordinario di Lingua e letteratura siriana, Università degli studi di Pisa).

<sup>14</sup> I fondi messi a disposizione dal progetto hanno coperto una parte delle digitalizzazioni. La differenza è stata coperta dalla Biblioteca stessa.

# Digitisation of Western Medieval Manuscripts at the British Library

**Claire Breay**  
*British Library*

The Medieval and Earlier Manuscripts Section at the British Library holds one of the richest collections of medieval and renaissance manuscripts in the world. These collections include over 25.000 western medieval manuscript books, over 3.000 Greek papyri and over 50.000 medieval charters and rolls. In recent years, the Medieval and Earlier Manuscripts Section of the Library has worked with Leipzig University Library, Saint Catherine's Monastery on Mount Sinai and the National Library of Russia to deliver the Codex Sinaiticus Project <http://www.codexsinaiticus.org>. This Website was first launched in July 2008 and updated in July 2009 to include full digitisation of Codex Sinaiticus, the fourth-century Greek biblical manuscript which includes the earliest text of the complete New Testament. The Website offers a new electronic transcription of the whole text of this heavily corrected manuscript, in which every word in the transcription is linked to the corresponding word in the images. This transcription was produced by a team based at the Institute for Textual Scholarship and Electronic Editing at the University of Birmingham. The Website also includes the results of a detailed page-by-page examination of all the physical characteristics of the manuscript undertaken by conservators in each location. In addition, the images captured for the Website have been reused to produce a full printed facsimile of the manuscript, to be published in January 2011.

Another digital project run by the Medieval and Earlier Manuscripts Section at the British Library is the Catalogue of Illuminated Manuscripts <http://www.bl.uk/catalogues/illuminatedmanuscripts> which was first released in July 2003 and which has been updated regularly since then with additional content so that it now provides a searchable database of illuminated and decorated manuscripts in the Arundel, Burney, Egerton, Hargrave, Harley, Henry Davis, Hirsch, King's, Lansdowne, Sloane, Royal, Stowe and Yates Thompson collections. The searchable catalogue records include selective digitisation of pages containing illumination and decoration. The latest items to have been added to the catalogue are illuminated manuscripts from the Royal collection, a selection of which will be displayed in a major temporary exhibition at the British Library opening in November 2011. Moving on from both of these projects, the Medieval and Earlier Manuscripts Section has been developing a longer-term plan to digitise the Library's western medieval manuscript collections in full. As a first step towards this goal, in 2009,

the Library initiated an 18-month pilot project, generously funded by a grant from the Stavros Niarchos Foundation, to digitise manuscripts written in Greek. The aim was to digitise completely 250 manuscripts, but by the end of the project 284 Greek manuscripts had been digitised in full. The new online resource, Digitised Manuscripts, was launched on 26 September 2010 and is freely available at <http://www.bl.uk/manuscripts>.

The 284 manuscripts currently available comprise over a quarter of the British Library's collection of around 1.000 Greek manuscript volumes which, together with the Library's collections of papyri, ostraca and Greek printed books, form part of one of the largest and most important resources for the study of Hellenic culture outside Greece. The Library's Greek manuscripts contain unique and outstandingly rich information for researchers working on the literature, history, science, religion, philosophy and art of the whole of the Eastern Mediterranean in the Hellenistic, early Christian, Byzantine and Ottoman eras. These manuscripts are in regular use by researchers in the Manuscripts Reading Room at the British Library in London. The 284 manuscripts selected for this project are representative of the broad range of manuscripts in Greek held by the Library and include some notable highlights of the collection, including:



– *The Theodore Psalter* (Add MS 19352). Produced in Constantinople in 1066, this highly illustrated manuscript of the Psalms is arguably the most significant surviving manuscript illuminated in Constantinople. One of the greatest treasures of Byzantine manuscript production, it is of pivotal importance for the understanding of Byzantine art. Made for Abbot Michael of the Studios monastery, it is named after its scribe and illuminator, the monk Theodore who produced 435 marginal illustrations that act as a commentary on the text of the Psalms.

Figure 1. *The Theodore Psalter*: BL Add. MS 19352, f. 91v



Figure 2. *Illuminated Gospels*: BL, Harley MS 1810, f. 261v

- *Illuminated Gospels* (Harley MS 1810). A late twelfth century gospel book which unusually integrates seventeen narrative images of the life of Christ and the saints into the Gospels. Whereas portraits of the evangelists became a traditional feature of copies of the Gospels in Greek, narrative images were much rarer.
- *Dialogues of Lucian* (Harley MS 5694). This early tenth century manuscript is the oldest surviving manuscript of the works of second-century author, Lucian. The text of the *Dialogues* is accompanied by marginal commentaries, or scholia, in the hand of the first owner of the manuscript, Arethas of Patrae, Archbishop of Caesarea from 902. They illustrate the deep interest of a prominent Byzantine churchman in classical antiquity and its pagan literature.
- *Babrius's fables* (Add MS 22087). The discovery of this manuscript on Mount Athos in 1842 gave rise to the first edition of Babrius's fables in 1844 and this manuscript remains the principal source for this text. It contains 123 Aesopic fables and was corrected by the great Byzantine scholar, Demetrius Triclinius.
- *Breviarium Historicum* (Add MS 19390, ff. 24-55). A late ninth-century manuscript of the history of the Byzantine Empire from the death of the Emperor Maurice in 602 to 713, by Nicephorus, Patriarch of Constantinople. Only one other manuscript of this history survives in the Vatican Library. These two manuscripts preserve a very rare attempt by a Byzantine author to write what would be accepted as proper history.

Prior to digitisation, the 284 manuscripts selected for the project were carefully assessed by one of the Library's conservators and a detailed risk assessment was conducted. Owing to the rarity and fragility of the manuscripts, particular care was taken in the handling of the material for the project; additional training for staff involved was provided and improved digitisation cradles were created. This ensured that any risk of damage to the objects during the digitisation process was minimised. Additionally, following an initial condition assessment, twenty-seven of the manuscripts selected received further conservation treatment to ensure that they could be safely digitised. All the manuscripts were also examined by a conservator after digitisation as a final check on each volume.

Digitisation of the selected manuscripts began in the British Library's existing imaging studios in August 2009. Over the course of the following ten months, approximately 100,000 digital images were captured as the 284 manuscripts in the pilot project were digitised in full.

Prior to this project, the selected manuscripts were amongst the least accessible online of the Library's western manuscripts. 250 of the volumes were drawn from the Additional manuscripts collection. These manuscripts, acquired in the late eighteenth and nineteenth centuries, had very summary existing catalogue records which were not suitable for reuse as content metadata. In tandem with the digitisation process, new catalogue descriptions for these 250 Additional Manuscripts were researched and written by a cataloguer recruited to work on this project. Modern catalogue descriptions for the remaining 34 Harley Manuscripts included in the pilot project had been published in 1999 and so were suitable for re-use as metadata. A review process took place to ensure that audiences accessing the manuscripts on-line were provided with high-quality, accurate information in a standardised format for all the manuscripts selected. As well as serving as searchable content metadata for the Digitised Manuscripts project, the catalogue records are also being migrated into the British Library's new Integrated Archives and Manuscripts System which is in the process of unifying catalogues for archival and manuscript material from across the Library.

Work on the Web presentation began in January 2010. It employed an existing Web infrastructure used at the Library for previous digitisation projects. An example of this is the Renaissance Festival Books Website: <http://www.bl.uk/treasures/festivalbooks/homepage.html>. This infrastructure enabled the Library to mount images (in three sizes – thumbnail, page view, and enlarged view, together with a ruler for close examination of details of all pages) with descriptive page-level metadata.

A comprehensive search facility enables Website users to perform searches of metadata to retrieve relevant manuscripts. Importantly, searches can be conducted using both the Roman and Greek alphabets. In addition to a quick search option, users can refine their searches by using a date-slider or by specifying key-

words, manuscript number, title, author, scribe, provenance or bibliography. Users can also browse by manuscript number, author, scribe or title.

In the manuscript viewer, users can move sequentially through the pages of the manuscript or can choose to go to a particular page. Images may be viewed as single pages, as a two-page open book, or as the recto and verso of a single leaf, and all images can be zoomed for examination at a high degree of magnification.

The project held two workshops, one in January 2010 for a group of academic researchers who all make extensive use of various types of Greek manuscripts in their work, and a second in May 2010 for an international group of digital humanities specialists. At the first workshop, the researchers attending reviewed several existing Websites which provide full digital coverage of medieval manuscripts to identify features that they found particularly helpful. This exercise provided valuable input into the design of the Digitised Manuscripts Website. Development of the functionality of the Website is on-going and we plan to incorporate feedback from the second workshop into the development of the site which in future will deliver a much wider range of manuscript material and allow greater user-interactivity.

The official launch of the Website took place on 26 September 2010 to coincide with the beginning of the new academic term. Press and publicity activity included global syndication of the story via Associated Press. The international media response to the launch was excellent, with over 250 articles reporting the launch of the resource and acknowledging the support of the Niarchos Foundation, within three days of the British Library's press announcement. In addition, the numerous international newspapers in North America and Australasia published stories about the resource.

In October 2010, both conservation assessments and digitisation started for the second phase of the Digitised Manuscripts project. Following the receipt of a second generous grant from the Stavros Niarchos Foundation, the Library will be able to add a further 250 Greek manuscripts to the Digitised Manuscripts Website by January 2012. Upon completion, the Library will have delivered full digitisation of over half of its collection of Greek manuscript books. In addition to this second Greek manuscripts project, the Library has also secured funding from an individual donor to allow us to digitise in full 140 medieval scientific and medical manuscripts. This project will begin early in 2011 and the content will be delivered on the Digitised Manuscripts Website in 2012. We hope to be able to offer an increasing number and range of manuscripts thereafter, subject to successful project fundraising.

# The ENRICH project: Towards a European digital manuscript library

**Matthew James Driscoll**

*The Arnamagnæan Institute; University of Copenhagen*

## Background

The idea of using computers to provide greater access to medieval manuscripts and other primary sources dates from the late 70s and early 80s, when a number of attempts were made to apply relational database technology to manuscript studies, in particular in the form of searchable electronic catalogues. Unfortunately – but understandably – these projects generally relied on locally developed or proprietary software, with all the problems for long-term maintenance and interoperability that entails. Moreover, each system tended also to have its own standards with regard to the nature, extent and organisation of information included, reflecting the lack of often even national standards for manuscript description at the time.

In the mid-Nineties the advent of Standard Generalized Markup Language (SGML) and the World Wide Web gave new impetus to work on electronic manuscript cataloguing. At the same time, developments in digital imaging meant that manuscript holding institutions could provide an unprecedented degree of access to their holdings. With the rise of large-scale digital collections came an increased awareness of the central importance of metadata standards.

In November 1996 a meeting was held at Studley Priory, near Oxford, organised by Peter Robinson of de Montfort University and Hope Mayo from the Mellon-funded EAMMS project (Electronic Access to Medieval Manuscripts) and attended by representatives from major manuscript holding institutions in Europe and the United States, together with experts on MARC, the Berkeley Finding Aids project, the TEI (Text Encoding Initiative) and Dublin Core. A year later there was a similar meeting at Columbia University in New York which brought together many of the participants in EAMMS, Digital Scriptorium (also funded by the Mellon Foundation) and several other manuscript-related projects. These meetings, both attended by the present writer, confirmed that there was indeed not only a widespread awareness of the need for an international standard for manuscript description, but also a fairly broad consensus as to what form that standard should take and what the appropriate technical means were to implement it, viz. something along the lines of the *Guidelines for Electronic Text Encoding and Interchange* developed by the TEI, an international and interdisciplinary standards project established in 1987 to develop, maintain and promulgate hardware – and

software – independent methods for encoding humanities data in electronic form<sup>1</sup>. In 1999 funding was obtained from the Telematics for Libraries section of the European Union Fourth Framework research programme for the establishment of the MASTER project (Manuscript Access through Standards for Electronic Records), whose goal was to define and implement a general purpose standard for the description of manuscript materials using TEI-conformant XML<sup>2</sup>. The project ran through 2001 and was, by the standards of many EU-funded projects, reasonably successful, in that the system it developed was actually adopted by many large-scale electronic cataloguing projects. Among the largest and most important of these is *Manuscriptorium*, a digital library of manuscripts and early printed books developed and maintained by the Czech National Library in Prague<sup>3</sup>. The most recent version of the TEI *Guidelines*, TEI P5<sup>4</sup>, released in November 2007, contains a major new chapter on manuscript description which is based largely on the work of the MASTER project and the TEI Medieval Manuscripts Description Work Group (TEI-MMSS), active between July 1998 and October 2000, which was headed by Consuelo Dutschke of the Rare Book and Manuscript Library, Columbia University, and Ambrogio Piazzoni of the Biblioteca Apostolica Vaticana. Although the work of these two groups proceeded largely in tandem, and despite an avowed intention that a single set of recommendations should emerge from them, there were, in the end, some significant discrepancies between the two proposed schemes. The MASTER project, for example, never finalised its discussion on seals before the end of the project period, while TEI-MMSS did, whereas MASTER developed quite sophisticated mechanisms for dealing with bibliographical and prosopographical data, an area largely untouched by the Work Group. In this sense the two schemes could be said to complement each other. There were, however, also discrepancies between the two which seemed to reflect a fundamental difference of opinion as to what the system should be used for and by whom. Thus TEI-MMSS, which consisted principally of librarians and cataloguers, seemed primarily concerned with the practicalities of manuscript cataloguing, and in particular with the accommodation of existing (legacy) data, while the MASTER project, which consisted principally of manuscript scholars and mark-up experts, seemed more interested in determining the underlying structure of manuscript descriptions in a more general, theoretical way. In order to resolve this is-

<sup>1</sup> For information on the TEI see <http://www.tei-c.org>.

<sup>2</sup> Principal project members were The Centre for Technology and the Arts at De Montfort University, Leicester (UK), Oxford University's Humanities Computing Unit (UK), Koninklijke Bibliotheek, Den Haag (NL), L'Institut de recherche et d'histoire des textes, Paris (FR), Národní knihovna české republiky, Praha (CZ) and Det Arnamagnæanske Institut, København (DK). Unfortunately, the MASTER website was not maintained after the end of the project, but a number of cached copies of MASTER-related documents can be found on <http://xml.coverpages.org/master.html>.

<sup>3</sup> <http://www.manuscriptorium.com>.

<sup>4</sup> *Guidelines for Electronic Text Encoding and Interchange*, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

sue, the TEI Council in 2002 appointed a special task force, led by the present writer, whose job it was to review the current state of TEI-based recommendations for the detailed description of manuscript materials and define a common subset of those recommendations adequate to the needs of the TEI community. Because the task force was able also to take into account the actual experience of the many electronic cataloguing projects then under way, the manuscript description module eventually incorporated into TEI P5 is not simply a common subset of the two schemes, but rather a significant improvement on both<sup>5</sup>.

## ENRICH

In 2007 the ENRICH project received funding under the eContent*plus* programme with the aim of extending *Manuscriptorium* to create seamless access to distributed information on manuscripts and early printed books throughout Europe, while at the same time upgrading the underlying metadata from Master*plus* (essentially the MASTER standard with added structural metadata) to TEI P5<sup>6</sup>. One of the project's central work packages, WP3, dealt with the "standardisation of shared metadata". Its goal was to ensure interoperability of the metadata used to describe all the shared resources by analysing the various standards used by different partners and ensuring their mapping to a single common format, which will be expressed in a way conformant with current standards.

The first thing that was done within the project was therefore to assess what differences there actually were between TEI P5 and Master*plus* and then resolve these differences. A wide sample (more than 1.000) of existing manuscript description records in many formats was reviewed, allowing the identification of a common core of practice. On the basis of this, a narrow subset of the TEI – which is designed to support a huge range of document types and encoding practices – was defined, including only those elements needed for the description and transcription of primary sources, as well as elements for linking these descriptions and transcriptions to digital images, where they exist.

Other constraints were added, for example, by pre-defining the contents of many attribute value lists rather than leaving them open, and making a number of attributes obligatory rather than optional.

<sup>5</sup> For more information on this process see my article *P5-MS: A general purpose tagset for manuscript description*, <http://www.digitalmedievalist.org/journal/2.1/driscoll/>.

<sup>6</sup> Partners in the ENRICH project were: Národní knihovna české republiky, Praha (CZ), AIP Beroun, s.r.o., Beroun (CZ), Oxford University Computing Services (UK), Centro per la comunicazione e l'integrazione dei media, Università degli Studi di Firenze (IT), Matematikos ir informatikos institutas, Vilnius (LT), SYSTRAN s.a., Paris (FR), Biblioteca Nacional de España, Madrid (ES), Biblioteca Nazionale Centrale di Firenze (IT), Vilniaus universiteto biblioteka (LT), Biblioteka Uniwersytecka we Wrocławiu (PL), Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík (IS), Universität zu Köln (DE), Monasterium Projekt, Diözese St. Pölten (AT), Landsbókasafn Íslands – Háskólabókasafn, Reykjavík (IS), Budapesti Mészaki és Gazdaságtudományi Egyetem (HU), Poznańskie Centrum Superkomputerowo-Sieciowe (PL) and Den Arnamagnæanske Samling, Nordisk Forskningsinstitut, Københavns Universitet (DK).

Let us look at one example. The TEI manuscript description module defines a number of specific elements designed to contain certain types of information. These are:

- <msIdentifier>: groups information uniquely identifying the manuscript, such as holding institution and shelfmark;
- <msContents>: provides an itemised list of the intellectual content of the manuscript, with transcriptions of rubrics, incipita, explicita etc., as well as primary bibliographic references;
- <physDesc>: groups information concerning all physical aspects of the manuscript, its material, size, format, script, decoration, binding, marginalia etc.;
- <history>: provides information on the history of the manuscript, its origin, provenance and acquisition by its current holding institution;
- <additional>: groups other information about the manuscript, in particular administrative information relating to its availability, custodial history, surrogates etc.

Within each of these, further specialised elements are defined. The <physDesc> element, for example, can contain elements for describing features such as the nature of the support, the dimensions of binding, leaves and written area, the foliation, pagination and columnation, the collation or quire structure, the layout of the page, the scripts used and identification of the hands, of known, as well as descriptions of illumination, decoration, paratextual features, musical notation etc. Use of all of these elements, apart from <msIdentifier>, is optional in the TEI, and often there is more than one possible way to provide the same information. For the purposes of the ENRICH project, however, it was decided to make a large number of elements and attributes obligatory, in order to ensure that all partners provided at least some basic types of information and encoded it in the same way. In order to indicate the nature of the support, for example, it was decided that the @material attribute on the <supportDesc> element should be compulsory, and that it must take one of the following values: “perg”, for parchment, “chart”, for paper, “mixed” or “unknown”. In this way, the support is given for every manuscript in the system, and in a way which is searchable regardless of the language in which the manuscript description is written<sup>7</sup>.

Synchronising ENRICH’s requirements with TEI P5 necessitated close collaboration with the TEI Council, which was revising the manuscript module at the same time. It was also important to work closely with “AiP Beroun”, the private firm who acted as technical co-ordinator for the project, to ensure that the Manuscriptorium platform would in fact be able to support the full complexity of TEI P5. Finally, it was necessary that a complete consensus among partners was reached.

<sup>7</sup> For more information on the relationship between the ENRICH project and TEI P5 see <http://enrich.manuscriptorium.com/index.php?q=node/9>.

The ENRICH standard was formally defined using TEI ODD (One Document Does it all) – the source format in which the *TEI Guidelines*, including the schema fragments and prose documentation, are written in a single XML document – which allows the automatic generation of schemata in DTD (Document Type Definition) and the RelaxNG (Regular Language for XML Next Generation) and W3C (World Wide Web Consortium) XML schema languages, as well as full documentation in a variety of languages (French, Italian, Spanish and English). The ENRICH standard has been tested in many different training contexts and a suite of training materials produced, covering the basic ideas of XML markup as well as the TEI modules for metadata, basic document structure, manuscript description and transcription, persons and places, facsimiles and non-standard writing systems<sup>8</sup>. A suite of XSLT (eXtensible Stylesheet Language Transformations) stylesheets and associated workflows – collectively known as the “ENRICH Garage Engine” – has also been developed for conversion from existing metadata formats such as EAD (Encoded Archival Description), MASTER and MARC (Machine Readable Cataloging), while the ENRICH “Gaiji Bank” is a tool for dealing with non-standard characters and glyphs, something which is often crucial for those working with manuscripts and other historical documents<sup>9</sup>.

In sum, ENRICH provides a system which facilitates both the lossless conversion of existing manuscript description data and the creation of completely new data. What is more, ENRICH can be used to produce the complete digital surrogate, comprising a collection of digital images of the manuscript, an associated TEI Header – the metadata component of any TEI document – containing a description of the manuscript, an encoded transcription of the manuscript’s text(s), optionally incorporating layers of scholarly interpretation and analysis, and an associated body of factual information about e.g. the persons, places, organisations and events related to the manuscript – and link all these components seamlessly together.

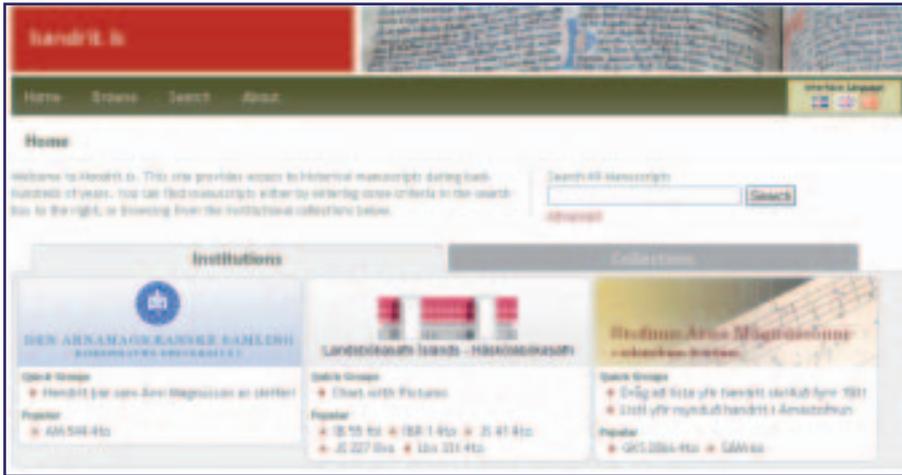
### Handrit.org

The way in which this works in practice can be seen from handrit.org., a digital library of Icelandic manuscripts, which is a collaborative effort by three partners in the ENRICH project, the Arnamagnæan Institute (Den Arnamagnæanske Samling) in Copenhagen, the Árni Magnússon Institute for Icelandic Studies (Stofnun Árna Magnússonar í íslenskum fræðum) in Reykjavík and the National and University Library of Iceland (Landsbókasafn Íslands – Háskólabókasafn).

Handrit.org was conceived as a central point of access for information about and analysis of the manuscripts in these three collections, which between them com-

<sup>8</sup> Links to all these documents and tools can be found on the OUCS website: <http://tei.oucs.ox.ac.uk/ENRICH/>.

<sup>9</sup> For the ENRICH Garage Engine see <http://dl.psnr.pl/software/EGE/>; for the Gaiji Bank, see <http://manuscriptorium.com/index.php?q=gaijibank>.



prise nearly 90% of the Icelandic manuscripts extant<sup>10</sup>. The system, which is currently in beta development stage, is based wholly on the native XML database eXist, with PHP used for the website front end. TEI-conformant XML manuscript descriptions are produced according to the ENRICH schema. These provide information on the manuscripts' contents, physical structure, origin and subsequent history. Controlled vocabularies are used to regulate content, typically through fixed lists of attribute values defined in taxonomies in the TEI Header or "hard wired" into the schema. One example of the former is the list of possible text-types available as values of the @class attribute on <msItem>. This list is based on collaborative work by Icelandic and Danish manuscript scholars and does not represent a "standard" as such, though it might well become one. In other cases existing international standards are used, and the value lists built into the schema. Extensive use is also made of authority files, e.g. for the names of persons, places and institutions, using the TEI elements <listPerson>, <listPlace> and <listOrg>, respectively. All proper names occurring in the individual manuscript descriptions are tagged using <name>, with a required @type attribute to indicate whether it is the name of a person, place or organisation/institution and a @key attribute which points to the relevant <person>, <place> or <org> element. In this way it is possible to search for manuscripts written at a certain time, in a certain place and containing certain types of texts. By combining these criteria with others relating, for example, to the social status of the scribes and owners and, say, manuscript format, a nuanced picture of Icelandic manuscript production and consumption over many centuries can be obtained.

<sup>10</sup> Other significant collections of Icelandic manuscripts are found in the Royal Library in Copenhagen, the Royal Library in Stockholm, Uppsala University Library, the British Library and the Bodleian Library in Oxford.

# “Internet Culturale. Cataloghi e Collezioni digitali delle biblioteche italiane”.

## La nuova versione del portale

**Laura Ciancio**

ICCU

Nel 1999 la Direzione generale per i beni librari e gli istituti culturali del MiBAC commissionava uno *Studio di fattibilità per la biblioteca digitale*, come quadro di riferimento alla formulazione di un progetto nazionale che consentisse al nostro paese di inserirsi in un contesto di iniziative analoghe internazionali. Lo studio e la riflessione organica seguita con la III Conferenza nazionale delle biblioteche del 2001<sup>1</sup>, evidenziarono la carenza di un coordinamento delle iniziative e di standard comuni di riferimento. In questi anni vengono poste le basi e definite le tappe successive in termini organizzativi e progettuali.

La cooperazione che derivò tra le diverse componenti, enti locali, istituti culturali, biblioteche e musei, portò alla elaborazione di una lunga serie di iniziative sul digitale finanziate nell'ambito del progetto BDI e NTC (Biblioteca digitale italiana e Network turistico culturale), presentato dalla Direzione generale per i beni librari e gli istituti culturali, approvato e cofinanziato dal Comitato dei ministri per la società dell'informazione (CMSI) nel 2003. Tra queste iniziative nel marzo 2005 veniva inaugurato il portale Internet Culturale, parte integrante del progetto BDI, destinato ad accogliere i risultati dei progetti. La componente NTC del portale enfatizzava la valorizzazione di quei beni che caratterizzano l'identità culturale del nostro paese, individuando alcuni temi maggiormente rappresentativi, musica, scienza, letteratura, modulando la loro presenza e vitalità nel contesto regionale e territoriale. La componente turistico culturale diventava decisiva per l'assegnazione del nome e ha condizionato la sua identità inducendo l'utente a percepirlo come un portale non strettamente legato al mondo bibliotecario.

Non è questa l'occasione per ripercorrere e analizzare la storia e le strategie che portarono alla pubblicazione di Internet Culturale; il confronto con le origini consente di spiegare, come nel contesto cambiato, sia stata necessaria una ridefinizione del suo ruolo e del suo posizionamento rispetto alle altre iniziative web del Ministero per i beni e le attività culturali (Mibac), sia nei confronti di portali aggregatori di informa-

<sup>1</sup> III Conferenza nazionale delle biblioteche *La Biblioteca digitale. Produzione, gestione e conservazione della memoria nell'era digitale*, Padova, Abbazia di S. Giustina, 14-16 febbraio 2001.

zione culturale come CulturalItalia, sia nei confronti delle stesse basi dati catalografiche specializzate gestite dall'ICCU. Nostra intenzione è presentare i risultati del progetto di reingegnerizzazione del portale appena concluso<sup>2</sup>.

### Nome e identità

La necessità di una ridefinizione della *mission* del portale era stata sollevata dallo *Studio sul riassetto, la riorganizzazione e la nuova prospettiva dei contenuti del portale Internet Culturale* commissionato dalla Direzione generale per i beni librari e gli istituti culturali al Dipartimento di italianistica e spettacolo dell'Università di Roma "La Sapienza"<sup>3</sup>, studio licenziato a dicembre 2008, base di partenza per la progettazione e restyling del portale.

Nelle riunioni del Gruppo di lavoro della Sapienza e del gruppo ICCU<sup>4</sup> che hanno preceduto la consegna dello studio, sono state valutate le analisi e le proposte formulate in rapporto alle linee d'azione della Direzione generale e dell'ICCU all'interno di un quadro di compatibilità economiche. Lo studio ha analizzato e definito i destinatari del portale, gli utenti dei servizi bibliografici, i cui *target* sono contenuti in tre macrocategorie: una utenza specialistica, una utenza scolastica e una utenza di base generica. Ha posto la questione del nome, ritenuto poco idoneo e non evocativo dei concetti di cataloghi bibliografici né di biblioteca digitale, non contribuendo in alcun modo alla definizione dei contenuti del portale. In quel contesto veniva ritenuta inopportuna la sua sostituzione, tuttavia, nella fase esecutiva del progetto è stato aggiunto il sottotitolo *Cataloghi e Collezioni digitali delle biblioteche italiane* al fine di evocarne le finalità. La definizione della *mission* e dei destinatari ha determinato l'impostazione degli interventi riguardo alla riorganizzazione dell'esistente, della presentazione e valorizzazione dei contenuti, ha determinato la progettazione dell'offerta di informazioni e di servizi.

### Obiettivi strategici

Obiettivo primario di Internet Culturale è promuovere la conoscenza del patrimonio librario italiano attraverso l'accesso ai cataloghi bibliografici e attraverso l'accesso al patrimonio digitale, e ancora valorizzare il patrimonio offrendo approfondimenti culturali sulle raccolte librarie.

L'accesso ai cataloghi doveva necessariamente esigere una differenza tra ciò che offre Internet Culturale e i cataloghi a cui attinge, e l'obiettivo prefissato in questo

<sup>2</sup> Il progetto è stato realizzato dal raggruppamento di imprese Liberologico srl capofila, Gruppo Meta e Centrica. Inizio delle attività ottobre 2009 e conclusione dicembre 2010. Coordinamento RTI: Paolo Bizzarri; Responsabile del procedimento: Laura Ciancio, ICCU ente appaltante.

<sup>3</sup> Gruppo di lavoro, coordinato da Giovanni Solimine (responsabile), Fabio Ciotti e Gianfranco Crupi. Ad esso collaborano Stefano Buscaglia, Chiara Faggiolani, Veronica Giannini, Giusi Vullo, e occasionalmente altri collaboratori.

<sup>4</sup> Gruppo dell'ICCU: Laura Ciancio, Gabriella Contardi, Gisella De Caro, Andrea Giuliano, Claudia Leoncini, Cristina Magliano, Massimo Menna, Giuliana Sgambati.

progetto è stato realizzare un accesso integrato ai cataloghi delle banche dati dell'ICCU, tuttavia, sviluppi futuri consentiranno di proporci come accesso unificato anche ad altre banche dati.

L'accesso alle collezioni digitali è la *mission* del portale, pari all'accesso integrato ai cataloghi, se non principale obiettivo. Altrettanto importante era dunque garantire la qualità dell'accesso alle collezioni e migliorare significativamente la fruizione delle risorse digitali.

L'approfondimento culturale attraverso prodotti multimediali, ha significato costruire metodi efficaci per l'accesso più diretto e intelligente ai contenuti culturali molteplici, esistenti sul portale, rispondendo in modo più preciso e qualificato alle richieste diversificate degli utenti.

L'ampliamento dell'utenza del portale è un altro obiettivo, intercettando i bisogni dell'utenza scolastica oltre che quella tipica delle biblioteche, infatti anche per questa esigenza, a breve, una selezione dei contenuti di Internet Culturale dalla valenza fortemente didattica, verrà distribuita sul portale InnovaScuola, iniziativa del Dipartimento per la digitalizzazione della pubblica amministrazione e l'innovazione tecnologica e del Ministero dell'istruzione, dell'università e della ricerca.

È stato poi necessario ai fini del rispetto della legge 9 gennaio 2004, n. 4 *Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici*, nota come "legge Stanca", rivedere i contenuti del portale esistenti perché fossero rimosse le inaccessibilità, per quanto possibile.

Tenuto conto delle risorse economiche gestite dal progetto e tenuto conto delle ridotte risorse umane a cui può fare ricorso la redazione del portale<sup>5</sup>, non è stato possibile ampliare in questo progetto l'offerta di una panoramica sul mondo dell'editoria da realizzare in collaborazione con il Centro per il libro del Mibac, come proposto dallo studio del Gruppo della Sapienza, ed ancora, individuare e promuovere nuovi approfondimenti culturali, compiti che la redazione potrà elaborare nei prossimi mesi chiedendo la collaborazione di tutte le biblioteche e istituzioni culturali che vorranno rendere disponibili anche sul portale Internet Culturale i propri prodotti digitali.

### Ricerca integrata – MetaIndice

La ricerca integrata, accesso rapido ai cataloghi, è collocata al centro della *home page* tra la sequenza di immagini della testata, che mostrano interni delle nostre biblioteche storiche alternati a pagine miniate, e lo spazio sottostante informativo. Una ricerca semplice, a campo unico, modalità familiare agli utenti web, dove, in alternativa all'impostazione di *default* su Cataloghi, è possibile interrogare separatamente la Biblioteca digitale o il Sito web. In fine il menu della Ricerca avanzata

<sup>5</sup> Redazione: Laura Ciancio (coordinatore), Valeria De Vita Cafasso, Daniela Napoletano, Elena Ravelli, Marco Scarbaci; per la Direzione generale biblioteche, istituti culturali e il diritto d'autore: Alfredo Esposito (coordinatore), Viviana Carini, Priscilla Sermonti.



per l'utente esperto. La ricerca su Cataloghi, intercetta i cataloghi SBN (Servizio Bibliotecario Nazionale), Edit16 (Censimento delle Edizioni italiane del XVI secolo), Manus (Censimento dei manoscritti delle biblioteche italiane), Biblioteca digitale, ReMI (Rete della musica italiana) e i materiali multimediali del portale (i Cataloghi storici al momento sono stati esclusi in quanto privi di metadati Mag). L'integrazione della ricerca bibliografica in una interfaccia unitaria è realizzata da un sistema complesso di indicizzazione delle diverse basi dati ed ha rappresentato l'intervento di maggiore complessità nell'ambito del progetto, dovuta in parte alla massa di dati (la sola estrazione dei records di SBN: più di 10 milioni) e alle differenti caratteristiche delle basi dati originali e loro formati di esportazione. È stato quindi definito un modello comune di dati adottando come set di elementi descrittivi del modello di dati le proprietà e gli schemi definiti formalmente nel Dublin Core Metadata Terms (DCMT). La mappatura delle singole basi dati è stata fornita dall'ICCU che a tal proposito ha costituito un gruppo di lavoro<sup>6</sup>. Quindi è stato creato un sistema di uniformazione dei contenuti sotto forma di metadati. Il reperimento dei dati viene garantito con *crawling* http, *crawling* su *file system*, e l'OAI-PMH 2.0. I dati delle diverse sorgenti vengono trasformati nel profilo comune mediante appositi *plug-in* o trasformatori XSL-T. Il Metaindice viene poi aggiorn-

<sup>6</sup> Gruppo di lavoro per la creazione della mappatura di MetaIndice: Tiziana Brunetti, Laura Ciancio (coordinatore), Gisella De Caro, Andrea Giuliano, Claudia Leoncini, Patrizia Martini, Cristina Mataloni, Massimo Menna, Donatella Roveri, Marco Scarbaci.

nato settimanalmente per i dati di SBN, mentre gli incrementi delle altre banche dati sono gestiti a richiesta; attualmente i sistemi di aggiornamento sono differenziati e organizzati in base alle specificità delle basi dati di provenienza.

Il motore di ricerca basato su software *open source* Lucene e SOLR crea degli indici specializzati per ciascuno dei campi presenti nel profilo comune. Su di essi compie le interrogazioni fornendo un punteggio (*rank*) di attinenza della risorsa rispetto alla richiesta. Tale punteggio si basa su algoritmi standard che fanno riferimento a tecniche di tipo statistico (TF-IDF, metrica coseno, ecc.) e può essere corredato da un meccanismo basato su tecniche di intelligenza artificiale di tipo CBR (*Case Based Reasoning*), in grado di tenere traccia delle scelte fatte in precedenza da utenti con profili simili. Per la gestione delle interrogazioni avanzate il motore di ricerca supporta gli operatori logici booleani tradizionali (AND; OR; NOT), la ricerca per frase; la ricerca per grafemi iniziali (parte iniziale di una parola). Inoltre, è supportata l'esplorazione ricorsiva di contenuti mediante l'utilizzo di tassonomie (Dewey browser). Il motore supporta, infine, la presenza di *thesaura* o ontologie per l'espansione semantica delle interrogazioni e l'individuazione automatica o semi-automatica di termini correlati.



Tale meccanismo è stato sperimentato per la gestione di sinonimi e pseudonimi negli autori. La presenza delle faccette, contenenti i metadati più significativi presenti nei documenti risultanti da una ricerca, permette di filtrare a posteriori e raffinare gli esiti di una interrogazione, combinando i criteri. In questo modo, si offre agli utenti la possibilità di non dover definire a priori il criterio da adottare ma di richiedere soltanto ciò che interessa (termini principali) per poi meglio focalizzarlo in un secondo momento sulla base dei risultati ottenuti.

L'erogazione centralizzata dei servizi di ricerca consente di dare uniformità all'esperienza dell'utente, la modalità di presentazione a faccette rende dinamici i risultati consentendo di elaborare e rinnovare la fruizione. Il risultato della ricerca di Cataloghi presenta schede sintetiche, da cui si accede alla scheda di dettaglio corrispondente della banca dati originale.

### **Biblioteca digitale**

L'Indice della Biblioteca digitale è specializzato per gli oggetti digitali presenti nelle Collezioni digitali descritti tramite metadati MAG. Si tratta di un archivio analogo a quello di Metaindice ma indipendente da esso. L'Indice è gestito e aggiornato a richiesta, mediante *harvesting* basato su protocollo OAI-PMH 2.0 verso le teche digitali come MagTeca dell'ICCU e le altre teche partner di Internet Culturale<sup>7</sup>.

In Biblioteca digitale la scheda sintetica dispone del *preview* della risorsa digitale, per consentire all'utente di accedere direttamente al visualizzatore, senza obbligatoriamente passare dalla scheda di dettaglio.

Nella ricerca di Biblioteca digitale i contenuti delle faccette sono quelli propri delle risorse digitali, perciò viene prospettata la descrizione della notizia e le informazioni relative alla risorsa digitale: l'agenzia che ha prodotto la risorsa, la collezione a cui appartiene la risorsa selezionata, il tipo di accessibilità alla risorsa, la completezza della digitalizzazione, il formato del digitale.

<sup>7</sup> Le teche digitali partner di *Internet Culturale* sono attualmente: Emeroteca della Biblioteca nazionale Braidense di Milano, Geoweb e MagTeca della Biblioteca nazionale Marciana di Venezia, Teca digitale della Biblioteca nazionale centrale di Firenze, Teca digitale del Museo Galileo di Firenze, Biblioteca italiana (BibIt) del Dipartimento di italianistica dell'Università degli studi di Roma "La Sapienza", Consorzio Baicr per Società geografica italiana, Biblioteca della Fondazione Basso e Biblioteca dell'Istituto Luigi Sturzo, Teca digitale dell'Istituzione Casa della musica di Parma, MagTeca dell'Istituto centrale per i beni sonori ed audiovisivi, MagTeca dell'ICCU per conto delle biblioteche: Conservatorio San Pietro a Majella di Napoli, Conservatorio Luigi Cherubini di Firenze, Fondazione Gioachino Rossini di Pesaro, Angelo Mai di Bergamo, Museo Donizettiano di Bergamo, Marucelliana di Firenze, Nazionale universitaria di Torino, Estense universitaria di Modena, Nazionale centrale di Roma, Archiginnasio e Museo internazionale della musica di Bologna, Provincia autonoma di Trento, Sisf – Sacro convento di Assisi, Abbazia di Montecassino, Oratoriana dei Girolamini di Napoli, Accademia filarmonica romana, Augusta di Perugia, Statale di Lucca, Istituto musicale Luigi Boccherini di Lucca, Società internazionale per lo studio del medioevo latino (SISMEL) di Firenze, Casanatense di Roma, Nazionale di Potenza, Accademia della Crusca di Firenze, Periodici preunitari, Museo nazionale del cinema di Torino e Cineteca nazionale di Roma.

Nella scheda di dettaglio delle risorse digitali presenti in Biblioteca digitale, oltre alla descrizione del contenuto, sono presenti alcuni strumenti volti a collegare la risorsa con altre che potrebbero presentare un interesse o una affinità con la risorsa stessa e quindi fornire all'utente la possibilità di esplorare la Biblioteca digitale secondo criteri trasversali. Lo scopo primario di tali strumenti è, infatti, quello di suggerire e stimolare o incuriosire l'utente verso contenuti della Biblioteca digitale di cui potrebbe non avere immediata percezione. A questo scopo, direttamente all'interno della scheda di dettaglio, i documenti che hanno relazioni di tipo madre – spoglio sono evidenziati e collegati (*link*) tra di loro. Inoltre, viene fatta una verifica se documenti con lo stesso identificativo (BID) sono presenti nelle altre banche dati (di Metaindice). In tal caso, compare un collegamento che permette la visualizzazione delle notizie relative a tali documenti correlati. È presente poi un meccanismo di “suggerimento” per la consultazione di altre risorse basato sul concetto di documenti simili. L'individuazione di tali documenti si basa sia su informazioni di tipo semantico che su algoritmi di tipo statistico che agiscono su alcuni campi di metadati al fine di individuare le similitudini tra le parole chiave presenti. È stata introdotta l'individuazione di similarità mediante l'utilizzo dei contenuti dei me-

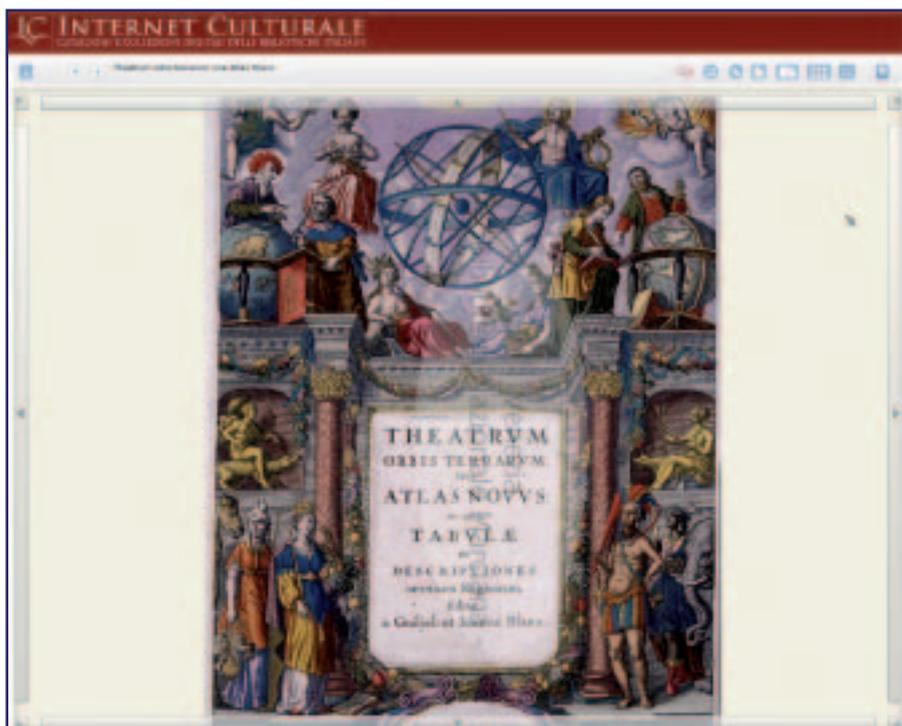


tadati *dc:title*, *dc:creator*, *dc:relation*, *dc:contributor*. Infine, sempre ai fini di fornire suggerimenti sulle risorse correlate, sono state integrate nel motore di ricerca tecniche di *clustering* (raggruppamento statistico non supervisionato di documenti) basate sul Soggetto della risorsa (se presente). Tali tecniche organizzano per similarità i documenti aventi lo stesso soggetto, mediante l'estrazione di descrittori (temi) di tipo semantico, l'individuazione su base statistica delle correlazioni, la selezione dei più significativi e la creazione dei raggruppamenti di documenti simili (*clusters*). Tali descrittori diventano quindi suggerimenti per possibili argomenti di ricerca (temi appunto) ed esplorazioni successive della Biblioteca digitale.

## Visualizzatore

Se la ricerca integrata nelle diverse banche dati rappresenta un servizio essenziale del portale Internet Culturale, l'accesso alle risorse digitali, rese disponibili dai vari progetti di digitalizzazione promossi nell'arco di dieci anni dalla Direzione generale per le biblioteche, gli istituti culturali ed il diritto d'autore e coprodotti con diverse istituzioni, ne costituisce il servizio qualificante.

Il visualizzatore che era in funzione sul portale presentava numerose criticità sia in termini di efficienza sia di funzionalità. Era necessario che il nuovo strumento garantisse la gestione dei metadati strutturali (sezione STRU del MAG) per navigare all'interno delle risorse digitali nel loro complesso (per esempio, le pagine di un li-



bro, le tracce di un album musicale). Era necessario uno strumento di ingrandimento per l'accesso a porzioni dell'immagine che fosse anche in grado di gestire risorse in formato testo.

Il nuovo visualizzatore supera queste criticità: sulla sinistra dello schermo vicino al bottone delle Informazioni è possibile visualizzare l'indice del volume come albero strutturato e attivo, quando valorizzato nel MAG, che consente all'utente di puntare direttamente ad una pagina del volume o a un brano da ascoltare.

Quando l'utente richiede di visualizzare uno degli oggetti digitali presenti nella Biblioteca digitale, il portale trasmette la richiesta a uno speciale componente, chiamato Multimedia Server (MMS). Il Multimedia server esamina la richiesta per l'oggetto digitale, e contatta la teca presso cui è presente l'oggetto digitale, ne estrae una versione a bassa qualità (comunque quella messa a disposizione della teca) e la rende immediatamente disponibile all'utente finale.

La presentazione è a pagina singola, ma è possibile la visualizzazione multi-pagina per selezionare tra i *preview* la pagina desiderata. Sono a disposizione le operazioni di *zuming*, di misurazione, di rotazione, di visualizzazione a schermo intero; se consentito dai detentori dei diritti è possibile scaricare una selezione o l'intera risorsa in formato pdf. Nel software è sviluppato un riproduttore di flussi audiovideo per il servizio di *media streaming* da parte del portale.



## Contenuti del portale

I due box vetrine centrali della *home page* riguardano Collezioni digitali e Temi, pagine della voce Esplora di menu. In Collezioni digitali sono elencate e descritte tutte le collezioni digitali indicizzate in Biblioteca digitale e in Metaindice, ciascuna scheda è corredata di fotografie e rinvio a scheda dell'istituto collegato, in qualche modo alla collezione, definito partner del portale. Nella voce di menu Esplora dopo le Collezioni trovano posto le categorie degli oggetti multimediali ereditati dal portale, con qualche riposizionamento: percorsi culturali, mostre, percorsi 3D, viaggi nel testo, autori e opere.

I Temi rappresentano una delle novità del portale. Al fine di un allargamento del *target* di utenza e di una ricaduta nel mondo della scuola, ci si è posti il problema di come valorizzare al meglio quei contenuti del portale appena citati. A cura della redazione questi materiali sono stati indicizzati e quando presenti, sono stati individuati come oggetti autonomi i contenuti sottostanti, elaborando metadati e arricchendo di soggetti e brevi descrizioni la lista dei risultati che, mappati, vengono ricercati in MetaIndice. Ancora, sono state utilizzate le categorie del Dewey, rappresentato dalla voce Temi nel menu Esplora, come metodologia di classificazione, per consentire all'utente di raggiungere attraverso quelle categorie gli argomenti sviluppati come autonomi contenuti in questo patrimonio di considerevole valore didattico.



Nella colonna di sinistra di ogni pagina è presente la voce Eventi e Novità, le cui notizie sono elaborate e gestite dalla sezione della redazione presso la Direzione generale delle biblioteche, degli istituti culturali e del diritto d'autore.

In conformità con le scelte operate, alcune voci non sono più presenti sul portale, rinviando l'utente ai siti istituzionali preposti alle informazioni di tipo istituzionale. Non si è voluto in alcun modo sostituirsi ai compiti che spettano ai servizi bibliotecari delle banche dati catalografiche, raggiungibili dall'utente in MetalIndice nel momento della selezione di una notizia, compreso il servizio di prestito. Dalla pagina Strumenti, dove è possibile accedere alle banche dati dei cataloghi, dell'anagrafe delle biblioteche, ai repertori, alla rivista «*DigItalia*», e ad una pagina di *link* ragionata sulle risorse della rete. Infine tutte le notizie contenute sul portale, compreso lo storico di Eventi e Novità, sono ricercabili dal comando ricerca specificando Sito IC.

### **Partner del portale e servizi di Internet Culturale**

I partner, a cui è dedicata una voce di menu, sono presentati con schede descrittive dell'istituzione accompagnate da fotografie. Si tratta degli istituti che a vario titolo hanno collaborato e collaborano con il portale: Regioni e Comuni cofinanziatori e promotori, enti di ricerca che forniscono analisi e collaborazioni scientifiche, consorzi digitali, istituti culturali, biblioteche con le proprie collezioni digitali. Nelle schede dei partner/istituzioni il link *Accedi al patrimonio* rinvia ai risultati di Biblioteca digitale, una risposta complessiva sul patrimonio digitale presente in Biblioteca digitale di quell'istituzione. Anche nelle schede delle collezioni, con *Accedi alla collezione* si ottiene il risultato della ricerca sulla sola collezione in Biblioteca digitale. Questa disposizione consente alle biblioteche e agli istituti partner del portale, attraverso i propri siti istituzionali di disporre di un *link* a questo risultato.

Questa funzione è stata realizzata per favorire la partecipazione del maggior numero di biblioteche al portale, le quali possono realizzare progetti di digitalizzazione devolvendo i costi del progetto alla sola produzione dei metadati e degli oggetti digitali, con notevoli risparmi nei costi di gestione, affidando i risultati al servizio realizzato dall'ICCU, con la teca digitale MagTeca<sup>8</sup>. I servizi che l'ICCU offre dal 2005 alle biblioteche partner del portale, sono di gestione gratuita delle risorse digitali, con l'archiviazione dei metadati secondo lo Schema MAG e delle risorse digitali a bassa e media risoluzione, garantendo nel tempo la conservazione e la disseminazione attraverso i servizi integrati di Internet Culturale.

<sup>8</sup> Il software si basa sul *framework* Fedora Commons 3.0, ed è stato realizzato da Inera srl, in base allo Studio di fattibilità rilasciato dalla Scuola normale superiore di Pisa, nel dicembre 2008.

## Nuovi partner di Internet Culturale

Affermando la propria identità culturale, come portale di accesso al patrimonio delle biblioteche italiane e punto di riferimento per quanti sono portatori di interessi nel mondo del libro, Internet Culturale in quanto aggregatore di contenuti digitali aspira ad accogliere i risultati di queste attività realizzate, oltre al progetto Biblioteca digitale italiana, da tutta la comunità bibliotecaria.

Chi aderisce può farlo disponendo di due modalità: le teche digitali potranno distribuire le risorse digitali colloquiando con i servizi di *harvesting* di Internet Culturale direttamente attraverso il protocollo OAI-PMH; le biblioteche, che realizzano progetti di digitalizzazione, affidando le proprie risorse digitali e metadati alla MagTeca dell'ICCU che trasmette con *harvesting*, protocollo OAI-PMH, ai servizi del portale.

Andranno concordate tra i servizi del portale e le istituzioni partner, attività tecniche di analisi e mappatura nel caso di banche dati catalografiche, attività tecniche di analisi della qualità dei metadati, sia per l'acquisizione nella MagTeca, sia per l'*harvesting* verso i servizi del portale, suggerendo linee guida sui requisiti tecnici da rispettare in sede di digitalizzazione, per favorire la ricerca e la presentazione dei contenuti. Le linee guida verranno pubblicate nella pagina del portale dedicata alle nuove adesioni.

Infine, vogliamo ricordare che chi aderisce al portale partecipa ad una comunità più ampia e di risonanza extra settoriale, vedendo garantita l'informazione dei propri dati nel contesto di CulturalItalia, e a livello internazionale sul portale Europea.

# I progetti di digitalizzazione del Museo Bodoniano di Parma

**Andrea De Pasquale**

*Museo Bodoniano*

Il Museo Bodoniano di Parma conserva una delle più ricche collezioni al mondo di materiali di varia natura (oggetti, volumi, documenti) pertinenti alle tecniche della fonderia di caratteri e alla tipografia nel periodo della stampa manuale. L'eccezionalità del materiale è anche data dal fatto che la totalità dei pezzi è relativa all'attività di un'unica azienda, quella gestita dal celebre tipografo Giambattista Bodoni (1740-1813) e proseguita dalla vedova Margherita fino alla sua morte (1841). Questa peculiarità permette lo studio, anche nei dettagli, della vita di un'officina tipografica d'*Ancien Régime*. Nello stesso tempo però la massa sterminata dei pezzi superstiti necessita, per poter essere compresa appieno, di essere fruita in maniera integrata, collegando i vari oggetti tra loro al fine di renderli anche comprensibili ad un pubblico non specialistico.

Spinti da tale esigenza e in vista del bicentenario della morte del Bodoni nel 2013, nel luglio del 2008 il Museo Bodoniano di Parma ha dato avvio ad una serie di lavori di digitalizzazione del materiale in esso conservato destinati a proseguire con gli anni. Tutto ciò è stato possibile grazie ai fondi della Biblioteca Palatina, che ospita nei suoi locali il Museo, e del Museo stesso. Le due istituzioni, oltre a promuovere la conoscenza di tale patrimonio e la sua maggiore visibilità, hanno come fine anche quello di garantire la tutela, permettendone infatti una lettura attraverso le rispettive immagini e quindi evitandone la manipolazione.

## Il materiale

### Materiale fusorio e tipografico

L'immenso patrimonio della fonderia dell'illustre tipografo, comprendente quasi 80.000 pezzi, è costituito da diverse casistiche di oggetti, in particolare punzoni, sciolti e in scatole lignee, matrici, con relativi strumenti per la fabbricazione e gli armadi per la conservazione, strumenti per la realizzazione della lega tipografica, forme e colini per la fusione per caratteri, pialle e vari strumenti per la rifinitura e il controllo di qualità dei caratteri, caratteri cinesi in legno, un compositoio e un vantaggio<sup>1</sup>.

<sup>1</sup> Andrea De Pasquale, *La fucina dei caratteri di Giambattista Bodoni*, Parma: Mup editore, 2010.

### L'archivio

Comprende le carte personali e di lavoro del Bodoni e della vedova, tra cui i ricchi carteggi, costituiti dalle lettere ricevute di corrispondenti e dalle minute. Ad esse si affiancano un'eccezionale raccolta di circa 200 manoscritti utilizzati per la composizione delle edizioni stampate<sup>2</sup> e alcuni documenti pertinenti ad altri personaggi legati ai coniugi Bodoni, gli amici e mecenati Felice e Nicolas De Azara, Giuseppe Bodoni, fratello di Giambattista, Antonio Comellini e Giuseppe Paganino, rispettivamente responsabili della fonderia e della stamperia di Bodoni dopo la sua morte, e le carte della gestione dell'eredità bodoniana, curata da Angelo Pezzana, direttore della Biblioteca Parmense (ora Palatina), nonché altri documenti giunti per acquisto sul mercato antiquario agli inizi del XX secolo.

### Le edizioni

Il Museo ha in consegna una raccolta di circa 800 edizioni stampate da Bodoni come direttore della Stamperia Reale e come gestore della sua tipografia privata, nonché diverse edizioni prodotte dalla vedova, spesso conservate con coperte editoriali dette "alla bodoniana" dal tipico colore arancione. Esse, insieme alle numerosissime altre ancora inserite nei fondi della Biblioteca Palatina, costituiscono la più ricca collezione al mondo di edizioni bodoniane. Si conservano anche più copie delle stesse edizioni con varianti testuali, su supporti diversi, e bozze corrette dei testi, fondamentali per comprendere il lavoro della stamperia.

Il Museo contiene anche diversi volumi appartenenti alla biblioteca personale del Bodoni, tra cui una straordinaria raccolta di campionari di caratteri delle principali fonderie europee<sup>3</sup>.

### Patrimonio storico-artistico e iconografico

Le collezioni comprendono anche una sessantina di oggetti, quali sculture, medaglie commemorative e lastre calcografiche usate nelle edizioni relative alla figura di Bodoni, e una serie di disegni e di incisioni relative a luoghi e personaggi in rapporto con la sua vita e la sua attività<sup>4</sup>.

## La campagna di digitalizzazione

### Materiale fusorio e tipografico

Il progetto ha previsto la digitalizzazione di tutti gli oggetti, realizzando immagini pertinenti ad oggetti singoli, ad insiemi omogenei (punzoni e matrici dello stesso alfabeto), ad oggetti seriali (scatole, lame, ecc.).

<sup>2</sup> Andrea De Pasquale, *I libri di lavoro di Giambattista Bodoni*, «L'Almanacco Bibliografico», 14 (giugno 2010), [http://centridiricerca.unicatt.it/creleb\\_Almanacco14.pdf](http://centridiricerca.unicatt.it/creleb_Almanacco14.pdf).

<sup>3</sup> Andrea De Pasquale, *La fucina dei caratteri* cit., p. 46-51.

<sup>4</sup> Per una rassegna su questi materiali cfr. Museo Bodoniano, *Catalogo del Museo Bodoniano di Parma*, compilazione e redazione di Angelo Ciavarella, Parma: Silva, 1968, spec. p. 60.

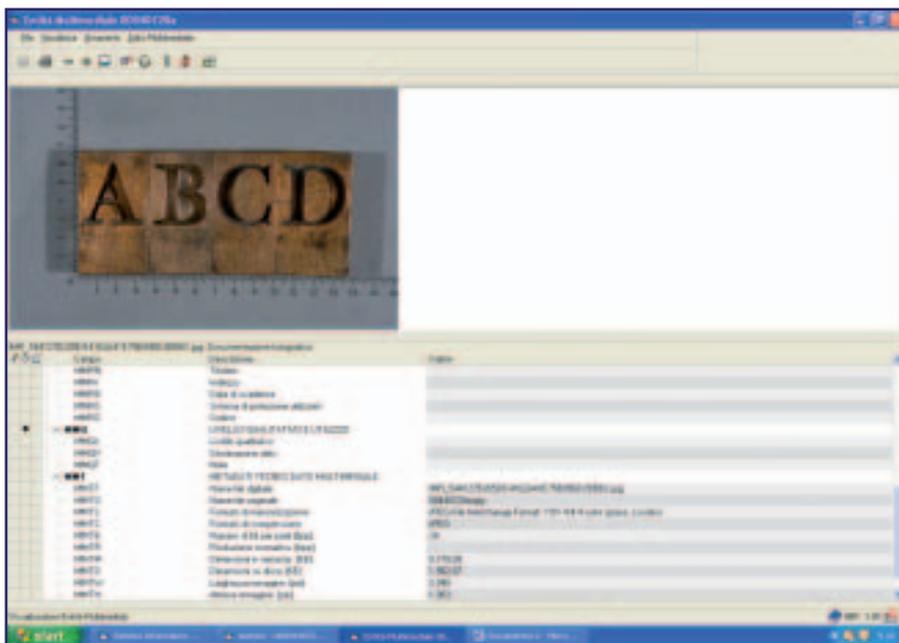


Figura 1a. Matrici per maiuscole latine tonde, non comprese nei caratteri, Perugia: schermata di scheda PST, campo FTA-documentazione fotografica, con metadati dell'immagine

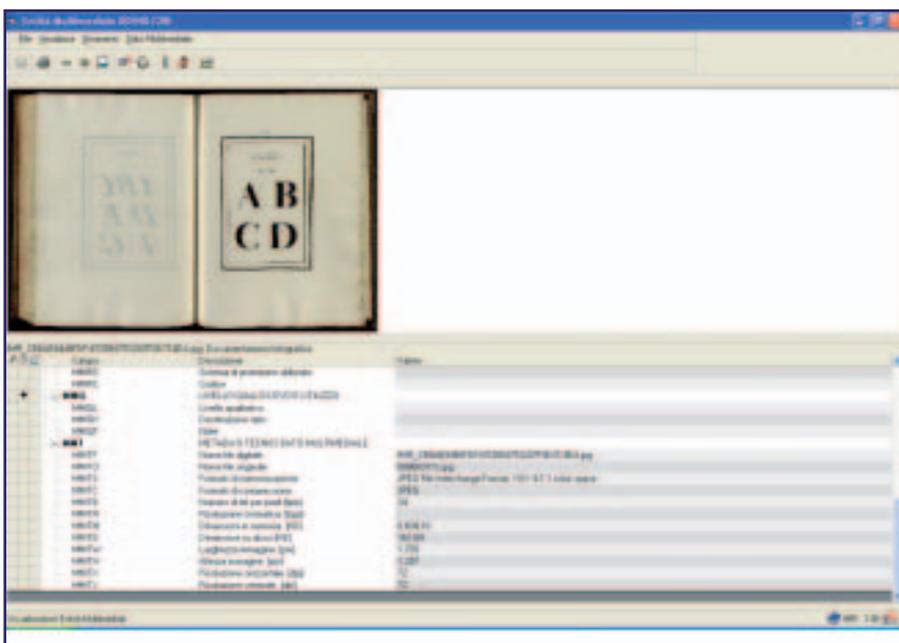


Figura 1b. G. Bodoni, Manuale Tipografico, 2 voll., Parma 1818 (vol. I, p. 254, alfabeto 108): schermata di scheda PST relativa all'oggetto di fig. 1.a, campo BIB-bibliografia, con metadati dell'immagine



zione, le più celebri e significative della produzione bodoniana. Inoltre spesso esse sono documentate attraverso l'affiancamento di copie in varianti, emissioni su supporti diversi (non solo carta, ma anche carta azzurra, pergamena e seta) e stati diversi. Si è scelto pertanto di digitalizzare, per tutte le edizioni in questione, la totalità delle varianti note, anche non esposte, conservate nei fondi della Biblioteca Palatina, previa un campagna specifica di catalogazione delle stesse.

### **L'archivio**

La prima campagna di digitalizzazione ha toccato esclusivamente le lettere ricevute da Giambattista Bodoni, pari a 7.631 lettere su bifogli per un totale di 18.255 immagini master.

Si è anche realizzata la digitalizzazione delle carte pertinenti alla Stamperia Reale, possedute dall'Archivio di stato di Parma, e appartenenti al fondo della Computisteria Farnesiana e Borbonica, conservate in 4 buste contenenti circa 5.000 carte, documenti fondamentali per studiare il funzionamento dell'officina diretta dal Bodoni.

### **Patrimonio storico-artistico e iconografico**

La digitalizzazione ha previsto la ripresa di 65 oggetti per un totale di 66 immagini master.

### **Tecniche di digitalizzazione**

#### **Materiale fusorio e tipografico e patrimonio storico-artistico e iconografico**

La digitalizzazione è stata realizzata nel rispetto delle norme dell'Istituto centrale per il catalogo e la documentazione (ICCD)<sup>5</sup>, attraverso l'utilizzo di macchina fotografica digitale, previo allestimento di un apposito set fotografico e l'apposizione di scala metrica.

Sono stati prodotti tre set di immagini aventi le seguenti caratteristiche: formato RAW/DNG: bit16; dimensione file: (cm 100 x 150) 72dpi rapportabili a (cm 24 x 36) con 300dpi; formato TIFF: bit8; dimensione file: (cm 100 x 150) 72dpi rapportabili a (cm 24 x 36) con 300dpi; formato JPG (ridotto in post produzione senza perdita di dati e utilizzato anche per la visualizzazione degli oggetti in abbinamento alle schede catalografiche): bit8; dimensione file: (cm 24 x 36) 72 dpi rapportabili a (cm 5,8 x 8,7) con 300dpi, compressione: 10.

<sup>5</sup> *La documentazione fotografica delle schede di catalogo: metodologie e tecniche di ripresa*, Roma: ICCD, 1998; *Normativa per l'acquisizione digitale delle immagini fotografiche*, Roma: ICCD, 1998; *Ambito di applicazione della "Normativa per l'acquisizione digitale delle immagini fotografiche" dell'ICCD*, febbraio 2005, tutti scaricabili alla url <http://www.iccd.beniculturali.it/index.php?it/115/standard-catalografici>.

### Edizioni e archivio

Per la digitalizzazione delle carte sono stati utilizzati scanner a planetario con ripresa dall'alto senza contatto con l'originale, dotati di piano basculante motorizzato. Espressamente progettati per la ripresa di materiale storico di pregio, essi dispongono di un sistema di illuminazione rigorosamente a luci fredde a LED bianchi e totalmente esente da emissioni di infrarossi e raggi UV.

Sono stati prodotti quattro set di immagini aventi le seguenti caratteristiche: formato TIFF, 24 bit, 300/400/600dpi a seconda del formato dei volumi, per quanto riguarda le edizioni e a 600 dpi per le carte d'archivio; formato JPG, 24 bit: alta risoluzione a 600dpi (100% del master), compressione 10; media risoluzione a 300dpi (50% del master), compressione 8; bassa risoluzione a 150dpi (25% del master), compressione 6.

### Catalogazione e metadati

#### Materiale fusorio e tipografico

Il lavoro di digitalizzazione si è affiancato a quello di catalogazione dei pezzi, primo esempio in Italia di schedatura di materiale di tale genere, attraverso gli standard descrittivi dell'ICCD.

È stata utilizzata la scheda PST (Patrimonio scientifico tecnologico) nella versione 3.0.1<sup>6</sup>, compilata a livello inventariale e solo per i campi di obbligatorietà assoluta, e software S.I.Ge.C. (Sistema informativo generale del catalogo), che prevede l'abbinamento alle schede descrittive delle corrispondenti immagini digitali, nonché la costruzione dei relativi metadati<sup>7</sup>.

Le schede delle matrici sono state collegate a quelle rispettive dei punzoni che presentano lo stesso alfabeto, attraverso il campo ROZ (riferimento orizzontale) e ad ogni scheda è stata abbinata, oltre alla/e immagine/i dell'oggetto o degli insiemi schedati, l'immagine corrispondente della pagina dei due documenti archivistici, il *Prontuario* del 1840 e l'*Inventario* del 1843, e, per i punzoni e le matrici, l'immagine della pagina del celebre *Manuale tipografico* del Bodoni pubblicato dalla vedova nel 1818, in modo da poter visualizzare la rappresentazione stampata dell'alfabeto corrispondente.

### Edizioni

Le edizioni sono state tutte catalogate in SBN a livello MED. Da esso sono stati scaricati i metadati descrittivi che sono stati completati, rispettando gli standard della Biblioteca Digitale Italiana (BDI) e della MAG-Teca, producendo XML secondo lo schema MAG 2.0.1<sup>8</sup>.

<sup>6</sup> La normativa è scaricabile dalla url <http://www.iccd.beniculturali.it/getFile.php?id=334>.

<sup>7</sup> Sul SGeC cfr. <http://www.iccd.beniculturali.it/index.php?it/118/sistema-informativo-generale-del-catalogo-sigec>.

<sup>8</sup> Su tale schema cfr. <http://www.iccu.sbn.it/genera.jsp?id=267>.





Figura 3. Piatta tipografica: schermata di scheda PST, campo FTA-documentazione fotografica, con metadati dell'immagine

## Prospettive

Il progetto di digitalizzazione è destinato ad ampliarsi, sulla base dei finanziamenti disponibili, a tutte le edizioni bodoniane, al fine di realizzare un completo repertorio digitale della produzione bodoniana, a tutte le minute delle lettere inviate e a tutte le restanti carte dell'archivio, ai manoscritti utilizzati per la composizione dei volumi, alle carte, spesso di riutilizzo, in cui è o era avvolto e conservato il materiale tipografico-fusorio. Inoltre sarà necessaria la redazione di tutti i metadati per gli oggetti digitali che ne sono privi.

Attualmente però il Museo sta lavorando alla messa a disposizione degli studiosi della sterminata quantità degli oggetti digitali già prodotti. A tal proposito è stata recentemente presentata una linea d'azione all'interno del piano lavori redatto dalla Direzione Regionale per i beni culturali e paesaggistici dell'Emilia Romagna nell'ambito del progetto nazionale MuseiD-Italia, finanziato con fondi europei, atto a rendere disponibile le informazioni possedute con la costruzione di un portale, che interrogherà, simultaneamente, anche dati pertinenti ad oggetti di nature diverse, e che le renderà consultabili anche attraverso Culturaitalia mediante protocollo OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting)<sup>11</sup>.

<sup>11</sup> Sul progetto e sulla specifiche cfr. <http://www.culturaitalia.it/pico/museiditalia/index.html>.

Si prevede infatti il *mapping* tra i metadati esistenti e quelli del profilo applicativo PICO (Portale Italiano della Cultura Online)<sup>12</sup>, la predisposizione di un apposito *server* dedicato, il popolamento dell'apposito *repository*, la verifica delle operazioni di *harvesting*.

Il progetto potrà essere ampliato con l'acquisizione di altre risorse digitali pertinenti al Bodoni ed è predisposto per accogliere contenuti di altra natura, quali video, atti ad illustrare le tecniche di fabbricazione di un libro antico e aspetti della biografia e della produzione del celebre tipografo.

<sup>12</sup> Sulle specifiche tecniche di adesione cfr. <http://www.culturaitalia.it/pico/FootMenu/documentazione/it/index.html>.

# Magazzini digitali: dal prototipo al servizio\*

Giovanni Bergamin – Maurizio Messina

Biblioteca nazionale centrale di Firenze – Biblioteca nazionale marciana di Venezia

## Introduzione<sup>1</sup>

Il progetto Magazzini digitali, avviato nel 2006 dalla Fondazione rinascimento digitale, dalla Biblioteca nazionale centrale di Firenze e dalla Biblioteca nazionale centrale di Roma si propone ora di mettere a regime un sistema per la conservazione permanente dei documenti elettronici pubblicati in Italia e diffusi tramite rete informatica, in attuazione della normativa sul deposito legale (L. 106/2004, d.p.r. 252/2006). Nella prima parte di questo contributo verrà descritta l'architettura tecnica del progetto, ma è bene tenere presente fin d'ora che la conservazione digitale, come è oramai ben noto, non si esaurisce solo in procedure di natura tecnologica. Le strategie volte ad evitare la perdita dei bit o a prevenire le dipendenze dall'hardware o dal software sono infatti solo una parte del problema. Vanno tenute nel conto dovuto le implicazioni economiche (la sostenibilità), la necessità di selezionare cosa è necessario conservare per le generazioni future, gli aspetti legali legati alla normativa sul diritto d'autore, la necessità della cooperazione fra le istituzioni titolari del deposito legale<sup>2</sup>. Alcuni di questi aspetti verranno esaminati nella seconda parte di questo contributo.

Ai fini del progetto, la conservazione digitale può essere definita come un servizio pubblico fornito da depositi *digitali affidabili* o *fidati* (*trusted* o *trustworthy digital repositories*) in grado di assicurare, per le risorse digitali depositate, la *leggibilità* a livello di bit da parte di una macchina (*viability*), la *interpretabilità* a livello di formato (per esempio: pdf, doc, ecc.) da parte di un elaboratore (*renderability*), l'*autenticità* (*authenticity*) intesa come identità e integrità dell'oggetto digitale, e la effettiva *disponibilità* (*availability*) per le comunità designate (comunità di riferimento, interessate all'uso di quelle risorse)<sup>3</sup>.

\* Viene qui riportato l'articolo di Giovanni Bergamin e Maurizio Messina corredato delle note non presenti nella precedente versione pubblicata nel numero 1 di *DigItalia* del 2010.

<sup>1</sup> Tutti i link sono stati controllati il 27 aprile 2010.

<sup>2</sup> Brian Lavoie – Lorcan Dempsey, *Thirteen ways of looking at... digital preservation*, «D-lib magazine», 10 (2004), 7/8, <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>.

<sup>3</sup> Queste definizioni sono basate su:

- *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*, [http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf) (per il concetto di "trusted digital repositories");
- Luciana Duranti, *Un quadro teorico per le politiche, le strategie e gli standards di conservazione digitale: la prospettiva concettuale di InterPARES*, «Bibliotime», 9 (2006), 1,

Il nome del progetto richiama intenzionalmente i magazzini delle biblioteche titolari del deposito legale. Come definito da uno storico progetto europeo sulla conservazione digitale (Networked European Deposit Library – NEDLIB<sup>4</sup>, svoltosi dal 1997 al 2000):

«For us, as memory organizations, this means we have to move from paper-based stacks to digital stacks».

Per molti aspetti i Magazzini digitali sono comparabili a quelli convenzionali: le risorse digitali devono essere conservate indefinitamente; i Magazzini digitali crescono man mano che si aggiungono nuove risorse; modifiche o cancellazioni di risorse non sono di norma possibili; è impossibile predefinire la frequenza d'uso delle risorse, alcune delle quali non saranno mai utilizzate, o lo saranno raramente<sup>5</sup>. Ed è interessante notare che nove anni più tardi, una ricerca su Google dei termini *Digital Stacks* restituisce la medesima espressione usata nel contesto della conservazione digitale:

«Digital stacks: rather than boxes, shelves, and climate controlled environments, digital information must be stored in containers, file systems, and secure servers»<sup>6</sup>.

### L'Architettura tecnologica

Lo scopo del progetto è stato quello di impiantare un'infrastruttura tecnologica con caratteristiche di "permanenza". Dando per assodato che i guasti o le disfunzioni dei vari componenti sono la norma piuttosto che l'eccezione<sup>7</sup>, l'infrastruttura è basata sulla replica dei dati (macchine differenti collocate in luoghi differenti) e su componenti hardware semplici e universalmente diffusi, non dipendenti dai produttori, e che possono essere sostituiti facilmente: in altre parole, semplici personal computer. Inoltre, l'infrastruttura non è dipendente da software proprietario ma si basa su sistemi operativi e servizi a codice sorgente aperto (*open source*).

<http://didattica.spbo.unibo.it/bibliotime/num-ix-1/duranti.htm> (per rendere valutabile l'autenticità di una risorsa digitale, un pubblico servizio deve essere in grado di stabilire la sua identità e di dimostrare la sua integrità);

- PREMIS 2.0, *PREsevation Metadata: Implementation Strategies*, 2008, <http://www.loc.gov/standards/premis/> (per i concetti di "viability, renderability, understandability, authenticity, identity");
- OAI: *Reference model for an Open Archival Information System*, ISO 14721:2003 (per il concetto di archivio e di comunità designata: «an organization that intends to preserve information for access and use by a designated community»).

<sup>4</sup> <http://nedlib.kb.nl/>.

<sup>5</sup> Jim Linden – Sean Martin – Richard Masters – Roderic Parker, *Technology Watch Report: The large-scale archival storage of digital objects*, 2005, <http://www.dpconline.org/docs/dpctw04-03.pdf>.

<sup>6</sup> <http://www.pedalspreservation.org/About/stacks.aspx>.

<sup>7</sup> Sanjay Ghemawat – Howard Gobioff – Shun-Tak Leung, *The Google file system*, 2003, <http://labs.google.com/papers/gfs-sosp2003.pdf>.

Attualmente un normale personal computer può facilmente immagazzinare fino ad 8 terabyte di dati, su 4 dischi da 2000 gigabyte, usando tecnologie SATA diffuse ed economiche<sup>8</sup>. La replica dei dati si basa su comuni programmi di utilità (*utility*) per la sincronizzazione dei dischi come *rsync*<sup>9</sup> e, per evitare dipendenze hardware come ad esempio le dipendenze da un determinato *disk controller*, non viene utilizzato il RAID (Redundant Array of Independent Disks)<sup>10</sup>.

Nel passaggio dal prototipo al servizio, inoltre, è stata modificata l'architettura tecnologica del *dark archive*. Il progetto originario prevedeva infatti l'uso di un sistema di *memorizzazione non in linea (offline storage)* basato su nastri di tipo LTO (Linear Tape-Open)<sup>11</sup>; successivamente si è deciso di utilizzare la medesima tecnologia già individuata per i due siti principali, cioè la *memorizzazione in linea (online storage)* su normali personal computer. L'espressione "in linea", comunque, non cambia la funzione del *dark archive*, che è quella di servire da archivio di sicurezza dei dati usabile in caso di *disaster recovery*<sup>12</sup>. I nastri LTO sono sicuramente una soluzione robusta ed affidabile introducono dipendenze e vincoli di natura tecnologica e gestionale (per esempio librerie automatizzate o robot). Per lo stesso motivo si è deciso di non usare un sistema HSM (Hierarchical Storage Management)<sup>13</sup>, in quanto le sue diverse implementazioni sono basate su sistemi proprietari.

La comparazione dei costi fra sistemi di memorizzazione in linea e non in linea non è facile: relativamente ai dischi SATA si può dire che il loro costo decresce in maniera proporzionale all'aumento della loro capacità, mentre è difficile stimare i costi complessivi di esercizio (il cosiddetto *total cost of ownership* – TCO) di una soluzione di memorizzazione basata su nastri<sup>14</sup>. Tenuto conto dei pro e dei contro si è deciso che la soluzione più conveniente fosse la memorizzazione in linea su semplici personal computer, facilmente sostituibili (*facilmente sostituibili* significa sostituibili con nessun impatto, o con un impatto trascurabile, sull'architettura complessiva).

L'unico inconveniente in questo approccio è effettivamente un problema ecologico: il consumo di energia e le emissioni di ossido di carbonio. Occorre però osservare che negli ultimi anni i cosiddetti computer ecologici (*green compu-*

<sup>8</sup> [http://it.wikipedia.org/wiki/Serial\\_ATA](http://it.wikipedia.org/wiki/Serial_ATA).

<sup>9</sup> «Rsync è un software per Unix che sincronizza file e cartelle da una posizione all'altra minimizzando il trasferimento di dati», <http://it.wikipedia.org/wiki/Rsync>.

<sup>10</sup> «RAID è un sistema informatico che usa un insieme di dischi rigidi per condividere o replicare le informazioni, combinandoli in una sola unità logica», <http://it.wikipedia.org/wiki/RAID>.

<sup>11</sup> [http://en.wikipedia.org/wiki/Linear\\_Tape-Open](http://en.wikipedia.org/wiki/Linear_Tape-Open).

<sup>12</sup> [http://www.webopedia.com/TERM/D/dark\\_archive.html](http://www.webopedia.com/TERM/D/dark_archive.html),  
[http://it.wikipedia.org/wiki/Disaster\\_recovery](http://it.wikipedia.org/wiki/Disaster_recovery).

<sup>13</sup> [http://en.wikipedia.org/wiki/Hierarchical\\_storage\\_management](http://en.wikipedia.org/wiki/Hierarchical_storage_management). HSM è una tecnica di gestione dello *storage* che sposta automaticamente i dati da apparati più veloci e di maggiore costo ad altri più economici e meno efficienti, a seconda dei diversi casi d'uso dei dati stessi.

<sup>14</sup> <http://digitalcuration.blogspot.com/2009/07/online-and-offline-storage-cost-and.html>.

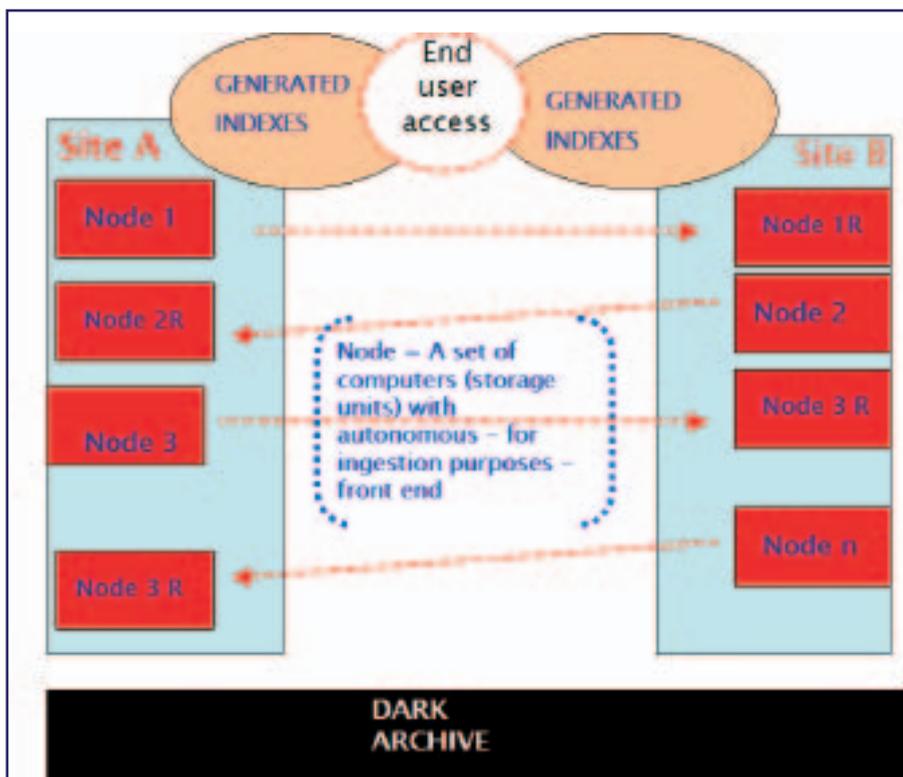


Figura 1. Modello dell'architettura tecnica di Magazzini digitali

ting)<sup>15</sup> stanno guadagnando quote crescenti e una diffusa coscienza da parte del mercato. Inoltre si sta sviluppando velocemente la tecnologia delle memorie a stato solido (SSD, Solid State Drive)<sup>16</sup>, prive di parti elettromeccaniche in movimento, e questo potrebbe ridurre significativamente nel prossimo futuro il consumo di energia delle apparecchiature di memorizzazione.

Grazie al finanziamento della DGBID (Direzione generale per le biblioteche, gli istituti culturali e il diritto d'autore), l'attuale prototipo di Magazzini digitali sta ora evolvendo in un servizio operativo basato su due siti principali di deposito, gestiti dalla Biblioteca nazionale centrale di Firenze e dalla Biblioteca nazionale centrale di Roma, e da un *dark archive* gestito dalla Biblioteca nazionale marciana di Venezia<sup>17</sup>. Naturalmente la Fondazione rinascimento digitale continuerà a supportare e promuovere il servizio.

<sup>15</sup> «This includes archival and backup data that would formerly have been saved on tape or other offline storage. The increase in online storage has increased power consumption. Reducing the power consumed by large storage arrays, while still providing the benefits of online storage, is a subject of ongoing research», [http://en.wikipedia.org/wiki/Green\\_computing](http://en.wikipedia.org/wiki/Green_computing).

<sup>16</sup> [http://en.wikipedia.org/wiki/Solid-state\\_drive](http://en.wikipedia.org/wiki/Solid-state_drive).

<sup>17</sup> Il servizio operativo è stato finanziato per tre anni alla fine del 2009.

Ciascun sito principale si compone di un insieme di *nodi* indipendenti ed autonomi. A sua volta ogni *nodo* è formato da un insieme di computer che gestiscono in autonomia le attività di acquisizione dei dati (*ingest*). Ciascun *nodo* su un determinato sito ha un corrispondente *nodo replica* sull'altro sito. Magazzini digitali non si basa quindi su un'architettura *sito principale/sito replica* (*master site mirror site*) dato che ciascun sito contiene, in maniera simmetrica, sia i nodi *master* che i nodi *mirror* (vedi figura 1). Ciascun file fisico è replicato due volte su computer diversi all'interno della stesso nodo. Anche il *dark archive* contiene due copie di ciascun file su due differenti computer. All'interno di Magazzini digitali, dunque, ciascun file fisico è replicato sei volte.

Collocare uno dei siti principali a Firenze sulle rive dell'Arno e l'altro a Venezia in Piazza San Marco, soggetta al noto fenomeno dell'"acqua alta", avrebbe costituito una seria minaccia per la sicurezza complessiva del servizio. Si è dunque deciso di collocare tutto l'hardware presso centri di elaborazione dati esterni, selezionati sulla base del possesso del requisito di base della certificazione secondo lo standard internazionale ISO 27001<sup>18</sup>. Ciascun istituto (BNCF, BNCR, BNM) selezionerà tre differenti centri di elaborazione dati posseduti e gestiti da tre differenti aziende, al fine di ridurre un possibile effetto "domino". Tali centri dovranno inoltre essere distanti l'uno dall'altro non meno di duecento chilometri, al fine di ridurre il rischio derivante da eventi catastrofici naturali. Questa architettura basata sullo standard ISO 27001 formerà la base per una specifica certificazione di Magazzini digitali come archivio digitale affidabile (o fidato). Durante la fase prototipale è stata condotta una sperimentazione con il sistema di autocertificazione DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)<sup>19</sup>, e si è tenuto conto anche delle specifiche di TRAC<sup>20</sup>.

### Metadati

Il sistema previsto per Magazzini digitali è in grado di accettare due tipi di file:

1. dati racchiusi in contenitori WARC: un contenitore WARC (ISO 28500) aggrega degli oggetti digitali per un agevole stoccaggio in un *file system* convenzionale<sup>21</sup>;

<sup>18</sup> ISO/IEC 27001:2005 «specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented Information Security Management System within the context of the organization's overall business risks».

<sup>19</sup> <http://www.repositoryaudit.eu/>.

<sup>20</sup> *Trustworthy Repositories Audit & Certification (TRAC)* cit.

<sup>21</sup> ISO 28500:2009 specifica il formato di un file WARC:

- «to store both the payload content and control information from mainstream Internet application layer protocols, such as the Hypertext Transfer Protocol (HTTP), Domain Name System (DNS), and File Transfer Protocol (FTP);

2. metadati racchiusi in contenitori MPEG21-DIDL: MPEG21-DIDL (ISO 21000) è un contenitore semplice ed agnostico adatto per l'archiviazione dei metadati (insiemi di metadati conformi a "schemi" differenti)<sup>22</sup>.

All'interno dell'architettura proposta Magazzini digitali deve affrontare il problema della gestione dei metadati, problema che è stato posto in termini di *modello lago/modello fiume*<sup>23</sup>: un archivio per la conservazione permanente non può basarsi su un *modello lago*, cioè sull'aggregazione di metadati conformi a pochi schemi<sup>24</sup> e alimentati da poche fonti principali, ma deve invece gestire la sedimentazione di metadati conformi a schemi che possono cambiare nel tempo e che sono alimentati da fonti molteplici; deve quindi basarsi sul *modello fiume*. In altre parole, in un archivio per la conservazione permanente dovranno convivere schemi di metadati che derivano, per usare il linguaggio di PREMIS (Preservation Metadata Implementation Strategies)<sup>25</sup>, da differenti *agent* (per esempio: *harvester* di meta-

- to store arbitrary metadata linked to other stored data (e.g. subject classifier, discovered language, encoding);
- to support data compression and maintain data record integrity;
- to store all control information from the harvesting protocol (e.g. request headers), not just response information;
- to store the results of data transformations linked to other stored data;
- to store a duplicate detection event linked to other stored data (to reduce storage in the presence of identical or substantially similar resources);
- to be extended without disruption to existing functionality;
- to support handling of overly long records by truncation or segmentation, where desired», [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=44717](http://www.iso.org/iso/catalogue_detail.htm?csnumber=44717).

<sup>22</sup> ISO/IEC 21000-2:2005 specifica:

- «Model: The Digital Item Declaration Model describes a set of abstract terms and concepts to form a useful model for defining Digital Items.
- Representation: The Digital Item Declaration Language (DIDL) is based upon the terms and concepts defined in the above model. It contains the normative description of the syntax and semantics of each of the DIDL elements, as represented in XML.
- Schema: Informative XML schemas illustrating complete grammars for representation of the DID in XML conforming to the normative representation.
- Detailed Examples: Illustrative (non-normative) examples of DIDL documents are provided to aid in understanding the use of the specification and its potential applications.

The ISO/IEC 21000 (MPEG-21) series of International Standards defines an open framework for multimedia delivery and consumption, with both the content creator and content consumer as focal points. The vision for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities.

This second part of MPEG-21 (ISO/IEC 21000-2:2005) specifies a uniform and flexible abstraction and interoperable representation for declaring the structure and makeup of Digital Items. A Digital Item Declaration (DID) involves specifying the resources, metadata, and their interrelationships for a Digital Item. A DID is done using the Digital Item Declaration Language (DIDL)», [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=41112](http://www.iso.org/iso/catalogue_detail.htm?csnumber=41112).

<sup>23</sup> <http://orweblog.oclc.org/archives/001754.html>.

<sup>24</sup> Il termine *Schema* è usato qui come definito in <http://www.w3.org/XML/Schema>: «XML Schemas express shared vocabularies and allow machines to carry out rules made by people».

<sup>25</sup> <http://www.loc.gov/standards/premis/>.

dati OAI-PMH – Open Archives Initiative-Protocol for Metadata Harvesting<sup>26</sup>, estrattori di metadati come JHOVE<sup>27</sup>, gli stessi bibliotecari, ecc.). Ogni schema è soggetto a cambiamenti nel corso del tempo e la sovrapposizione, a livello semantico, di elementi appartenenti a schemi diversi sarà probabilmente la norma piuttosto che l'eccezione.

Dal momento che i metadati sono l'unico mezzo per controllare i dati è essenziale avere il controllo dei metadati al fine di evitare il rischio di un "modello Babele". Attualmente si sta lavorando su questo, tenendo conto che non sembrano ancora esserci strumenti consolidati e disponibili. Ci sono però delle linee di sviluppo interessanti: tavole di corrispondenza (*crosswalks*) come *Morfrom*<sup>28</sup>, un *web service* dimostrativo di OCLC (On-line Computer Library Center), relativo a dati bibliografici, o i piani di sviluppo di *Dspace*<sup>29</sup> che dovrebbe implementare i risultati del progetto SIMILE (Semantic Interoperability of Metadata and Information in unLike Environments)<sup>30</sup>, un progetto di ricerca del Massachusetts Institute of Technology (MIT) e di HP che sta studiando come supportare schemi di metadati differenti usando RDF (Resource Description Framework)<sup>31</sup>.

## Il Modello giuridico e dei servizi

La seconda parte di questo contributo è relativa agli aspetti giuridici e agli accordi che sottostanno al progetto nonché al modello dei servizi.

La più recente normativa italiana sul deposito legale (L. 106/2004, D.P.R. 252/2006) prevede un periodo di sperimentazione del deposito legale su base volontaria dei documenti elettronici, definiti dalla legge «documenti diffusi tramite rete informatica»<sup>32</sup>. Tale normativa può essere considerata come la massima fonte di un impegno formalmente affidato alle biblioteche nazionali depositarie di costituire il nucleo di una rete nazionale per la conservazione digitale che, sulla base dell'esito della sperimentazione o limitatamente a specifiche tipologie documentarie, potrebbe comprendere anche le risorse elettroniche prodotte in altri domini, diversi da quelli in cui operano le biblioteche. Come è noto l'affidamento formale del "compito" di conservare le risorse è uno dei prerequisiti di un archivio digitale affidabile (o fidato)<sup>33</sup>.

La sperimentazione è finanziata dal Mibac, Direzione generale per le biblioteche, gli Istituti culturali e il Diritto d'autore, con il supporto umano, organizzativo e finanziario della Fondazione rinascimento digitale. Come precedentemente detto, viene condotta dalle BNCF e dalla BNCR, che operano come siti principali per l'ac-

<sup>26</sup> <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

<sup>27</sup> <http://hul.harvard.edu/jhove/index.html>.

<sup>28</sup> <http://journal.code4lib.org/articles/54>.

<sup>29</sup> <http://www.dspace.org/>, in particolare: <http://www.dspace.org/faq/FAQ.html>.

<sup>30</sup> <http://simile.mit.edu/>.

<sup>31</sup> [http://it.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://it.wikipedia.org/wiki/Resource_Description_Framework).

<sup>32</sup> L. 106/2004, art. 4.

<sup>33</sup> *Trustworthy Repositories Audit & Certification (TRAC)* cit.

cesso e la conservazione delle risorse elettroniche, e dalla BNM che gestisce il *dark archive* fuori linea, non accessibile al pubblico, per la ridondanza dei dati. I tre principali obiettivi della sperimentazione sono i seguenti:

1. implementare un modello organizzativo adatto a costituire gli archivi, nazionale e regionale, della produzione editoriale elettronica, come previsto dalla legge, e tale da poter essere esteso su scala più ampia;
2. implementare un modello di servizio tale da bilanciare gli interessi dei detentori dei diritti della protezione dei contenuti con quelli degli utenti finali all'accesso alle risorse;
3. implementare un sistema tale da assicurare l'accesso e la conservazione permanente dei contenuti digitali, e la loro autenticità (identità ed integrità).

Al fine di raggiungere tali obiettivi e di bilanciare i diversi interessi dei vari soggetti coinvolti (*stakeholder*) sono necessari degli accordi specifici:

1. un accordo fra le tre biblioteche nazionali e la Fondazione rinascimento digitale per definire le responsabilità ed i ruoli di ciascuna istituzione dai diversi punti di vista, scientifico, tecnico, operativo e finanziario e per istituire un Comitato di coordinamento per tutte le attività di gestione, monitoraggio e valutazione dei risultati. Compito del Comitato sarà anche quello di definire un piano per la sostenibilità finanziaria del progetto dopo i 36 mesi di sperimentazione; l'accordo, sotto forma di lettera d'intenti, è stato firmato il 19 gennaio 2010;
2. un accordo fra le tre biblioteche nazionali e ciascun editore elettronico (o ciascun fornitore di contenuti digitali) che parteciperà alla sperimentazione, relativo all'accesso e all'uso delle risorse digitali oggetto di deposito legale, tale da configurare un modello dei servizi. La normativa corrente (art. 38, D.P.R. 252/2006) prevede un accesso libero per via telematica ai documenti soggetti a deposito legale che siano in origine liberamente accessibili in rete, e un accesso limitato esclusivamente a utenti registrati che accedono da postazioni situate all'interno degli istituti depositari per quei documenti il cui accesso è originariamente soggetto a licenze o condizioni particolari. In ambedue i casi l'accesso deve avvenire nel rispetto delle norme sul diritto d'autore e sui diritti connessi. L'accordo, oramai concluso, prevede i seguenti punti:
  - BNCF e BNCR effettueranno periodicamente la raccolta (*harvesting*) dei documenti elettronici concordati con gli editori (l'*harvesting* è la modalità più semplice ed economica di alimentare l'archivio, anche dal punto di vista degli editori, a condizione che sia rispettata la normativa sul diritto d'autore);
  - nel caso di documenti accessibili su licenza, l'editore fornirà alle biblioteche le necessarie autorizzazioni, e verranno concordati i formati dei file (WARC etc.);
  - i documenti verranno immagazzinati in copie multiple (minimo 6 copie) in BNCF e BNCR, ed offline in BNM;

- le biblioteche saranno autorizzate a depositare i documenti presso *data center* esterni, certificati ISO 27001;
- gli archivi digitali saranno conformi ad OAIS (ISO 14721-2003) e saranno certificati come affidabili;
- BNCf, BNCr e BNM assicureranno l'accessibilità e la conservazione permanente dei documenti depositati, e manterranno traccia di qualunque loro modifica, fornendo un rendiconto trimestrale all'editore;
- BNCf, BNCr e BNM saranno autorizzate ad effettuare tutte le operazioni necessarie al mantenimento dell'accessibilità e della conservazione permanente dei documenti depositati (duplicazioni, migrazioni, etc.);
- i documenti depositati soggetti a licenza saranno resi consultabili solo da utenti registrati su postazioni multiple prive di stampanti, porte USB etc. poste sulle reti locali di BNCf e BNCr; tutte le operazioni svolte dagli utenti saranno tracciate, secondo le normative vigenti;
- la stampa o il *download* di file sarà soggetto a specifici accordi, e sarà previsto ove necessario un sistema di remunerazione del diritto d'autore (es. per i documenti protetti non disponibili sul sito dell'editore);
- l'accessibilità e la consultazione dei documenti depositati sarà consentita anche alle biblioteche regionali di deposito, con le stesse modalità, ma limitatamente ai documenti prodotti dagli editori la cui sede si trova nella stessa regione della biblioteca regionale di deposito.

Inoltre, al fine di estendere la base della sperimentazione, il progetto prevede il deposito anche dei seguenti tipi di risorse elettroniche, da regolare anch'esso tramite specifici accordi:

1. risorse digitali native prodotte dalle università e soggette anch'esse a deposito, con particolare riferimento alle tesi di dottorato;
2. risorse digitali risultanti dai progetti di digitalizzazione di materiali analogici finanziati dalla Biblioteca Digitale Italiana<sup>34</sup>, soprattutto nell'ambito delle istituzioni della memoria e limitatamente alle copie master.

Per quanto riguarda il primo di questi punti è stato attivato sperimentalmente il sito<sup>35</sup> del deposito legale, in cui compare una pagina informativa sulle procedure di deposito legale delle tesi di dottorato presso le biblioteche nazionali centrali.

Il servizio è stato realizzato e sperimentato in collaborazione con il Gruppo Open Access della CRUI (Conferenza dei Rettori delle Università italiane)<sup>36</sup>.

<sup>34</sup> <http://www.iccu.sbn.it/genera.jsp?s=18&l=it>.

<sup>35</sup> <http://www.depositolegale.it/oai.html>.

<sup>36</sup> La raccolta automatica (*harvesting*) dei dati e dei metadati delle tesi di dottorato di ricerca ai fini del deposito legale è prevista dalla Circolare MiUR n. 1746 del 20 luglio 2007.

La procedura di raccolta automatica (*harvesting*) delle tesi consente a «tutte le Università italiane che raccolgono le tesi di dottorato in formato digitale in un archivio aperto, secondo le raccomandazioni contenute nelle *Linee guida per il deposito delle tesi di dottorato negli archivi aperti*<sup>37</sup> approvate dalla Conferenza dei Rettori delle Università Italiane del 2007, di ottemperare agli obblighi di legge (senza ricorrere all'invio di documentazione cartacea)». Sono previste anche la raccolta e la conservazione delle tesi soggette ad embargo, che saranno consultabili solo all'interno delle reti locali delle biblioteche nazionali depositarie su postazioni prive di apparecchiature periferiche. La pagina informativa precisa inoltre:

1. il formato preferito per il deposito ovvero il PDF(A)<sup>38</sup>;
2. raccomandazioni per i metadati esposti dai *repository* delle Università mediante il protocollo OAI-PMH per facilitare l'interoperabilità sintattica e semantica;
3. specifiche tecniche per la raccolta di tesi digitali suddivise in più file, e indicazioni per la configurazione di EPrints3 e DSpace 1.5.

Alla sperimentazione della procedura di deposito legale delle tesi in formato digitale via *harvesting* automatico hanno partecipato l'Alma Mater Studiorum – Università di Bologna, l'Università Federico II di Napoli e l'Università di Trieste. Hanno contribuito alla definizione delle specifiche tecniche l'Alma Mater Studiorum – Università di Bologna per il software EPrints e l'Università di Trieste per il software DSpace. Hanno successivamente aderito alla sperimentazione la LUISS (Libera università internazionale degli studi sociali Guido Carli di Roma), l'Università di Parma, l'Università Cattolica di Milano, l'Università degli studi di Milano-Bicocca e l'Università di Venezia<sup>39</sup>.

L'ultimo aspetto da affrontare brevemente è relativo alla sostenibilità finanziaria del progetto Magazzini digitali: come è noto, l'accesso ai periodici elettronici è di norma soggetto ad una licenza. Una tipica clausola di tali licenze riguarda l'"accesso perpetuo" ai contenuti, che l'editore si impegna a garantire. Si tratta di una clausola di grande importanza sia per le biblioteche che per i loro utenti, e costituisce l'unico modo in cui le biblioteche possono garantire nel tempo la disponibilità di contenuti per i quali hanno sostenuto dei costi. Allo stesso tempo è una clausola che può essere rispettata solo attraverso la predisposizione di un'infrastruttura tecnica ed organizzativa dedicata (o terza), cioè un archivio digitale affidabile (o fidato); un archivio che è improbabile che gli editori abbiano interesse a gestire. Questo tipo di servizio potrebbe dunque essere affidato alla rete delle biblioteche di deposito legale, ed il suo costo potrebbe essere parte delle negoziazioni con gli editori per le licenze elettroniche<sup>40</sup>.

<sup>37</sup> <http://www.cruil.it/HomePage.aspx?ref=1149#>.

<sup>38</sup> [http://www.iso.org/iso/catalogue\\_detail?csnumber=38920](http://www.iso.org/iso/catalogue_detail?csnumber=38920).

<sup>39</sup> Alla data di chiusura di questo articolo (18 maggio 2010).

<sup>40</sup> Terry Morrow – Neil Beagrie – Maggie Jones – Julia Chruszcz, *A comparative study of e-journals archiving solutions: A JISC funded investigation: Final report*, 2008, <http://www.slainte.org.uk/news/archive/0805/jiscejournalreport.pdf>.



## **Documenti e discussioni**



# Delivering Content to Europeana in Practice: The ATHENA Harvesting Format LIDO\*

**Regine Stein**

*Philipps-Universität Marburg*

## Introduction

**A**THENA (Access to cultural heritage networks across Europe) provides content to Europeana by establishing a mechanism for harvesting museum holdings into Europeana. A major goal of the project is to develop an infrastructure that enables semantic interoperability with Europeana while preserving museum object specifics. To comply with this requirement, ATHENA has put particular focus on the choice of a format for content delivery that would be able to express the variety of museum's information appropriately. While the practical harvesting of data is an ongoing process and experience is growing, this article provides together with the actual presentation of the ATHENA harvesting format LIDO, some preliminary conclusions derived from the project.

## ATHENA's choice of a data model: ESE versus LIDO

The data model currently used in the Europeana prototype, ESE, is based on the Dublin Core metadata format. Although initially created strictly for the description of Web resources, Dublin Core has become the most common format in cultural heritage service environments. However, the ESE model is not considered as appropriate within the museum community: museum metadata is "flatten

out", with most of the data going into a limited subset of elements. For example, a number of different persons and institutions are usually associated with a museum object: the creator or finder of an object, important persons who have used it, the museum currently holding it, previous owners, and so on. All this qualified information is lost in the ESE format. Moreover, the lack of structure that allows elements to be grouped according to their semantic content leads to substantial information loss. A particular problem is the fact that Dublin Core does not allow information about the object itself and its digital surrogate to be clearly differentiated – the creator of the object appears in the same field than the photographer of its image.

Consequently, the ATHENA workpackage on metadata formats, following a best practice report on metadata formats used by the partners, came to the conclusion that a more appropriate data model for museum information should be used. Since the LIDO development already underway was primarily an effort to harmonize the two existing harvesting formats CDWA Lite and museumdat into one single schema, ATHENA decided to join the LIDO initiative and support further development that would subsequently integrate SPECTRUM requirements into the schema. Thus LIDO was

\* This article is a shortened version of the full paper *Sharing Museum Information: Theory or Practice – A European Experience* given at the CIDOC 2010 conference in Shanghai, November 10th 2010. For the full paper refer to [http://cidoc.meta.se/2010/full\\_papers/stein.pdf](http://cidoc.meta.se/2010/full_papers/stein.pdf).

chosen and further developed as the metadata format for the delivery of museum content through ATHENA to Europeana.

### The LIDO format

LIDO is an XML schema intended for delivering metadata, for use in a variety of online services, from an organization’s online collections database to portals of aggregated resources, as well as exposing, sharing and connecting data on the Web. The strength of LIDO lies in its ability to support the full range of descriptive information about museum objects; it can be used for all kinds of object, e.g. art, cultural, technology and natural science. Moreover, it supports multilingual portal environments.

LIDO defines 14 groups of information of which just three are mandatory. This allows for the widest and most comprehensive range of information possible. Organizations can decide on how rich – or how light – they want their contributed metadata records to be.

The schema consists of a nested set of “wrapper” and “set” elements, many of them repeatable, which organizes information about an object into a tree-like structure. This allows any degree of detail to be recorded in a logically correct, semantically coherent way. An important part of its design is the concept of events, taken from the CIDOC CRM. Information about actors, dates and places related to a museum object is

mediated through an event: the creation, collection, and use of an object are seen as events occurring during the object’s lifecycle. An exception is events that are depicted or referred to directly, considered as subject matter.

Another important construction principle is the distinction between indexing information that is optimized for searching and retrieval, and display information that is optimized for online presentation. Each information unit contains distinct sub-elements for indexing and display. The structural elements of LIDO contain “data elements” which hold actual data *values*. LIDO also allows the recording of information about data *sources* (e.g. in a book) and references to controlled terminology (e.g. the identification code for a term in a thesaurus). Conceptually the information in a LIDO record is organized in 7 areas, of which 4 have descriptive and 3 an administrative character:

The descriptive information section holds:

- object classification information such as object type and other classifications;
- object identification information such as titles, inscriptions, repository information, descriptions, and measurements;
- event information about events where the object was present or in which it participated, such as creation, modification, acquisi-



Figure 1. LIDO overview

tion, finding, or use. This section holds a number of sub-elements including event type and name, participating actors, cultures involved, date and place information as well as materials and techniques used (typically in the creation/production event);

- relation information links to related objects, but also to the subject – that is the content of a work: what is depicted in or by a work or what the work is about.

The administrative information section holds:

- rights associated with the object;
- record information about the source providing the metadata;
- resource information, in particular about digital resources being supplied to the service environment for representing an object online.

The result of a joint effort of several international key institutions and groups dealing with museum documentation standards, e.g. the CDWA, museumdat, SPECTRUM and CIDOC CRM communities, the release of LIDO v. 1.0 during this year's CIDOC conference can be seen as a clear reward to the community. It provides a single, common schema for contributing content to cultural heritage repositories. This enables museums and other content providers, using different data structures and software systems, to express and deliver a wide variety of information in a standardized and machine-readable format. Furthermore, this information can easily be accessed, harvested and recontextualized by semantic-aware services. Apart from the exciting promise of new applications, LIDO promises time – and cost – savings for museums interchanging object information in different daily work contexts.

### The ATHENA mapping and ingestion process

Now after this insight into the richness and opportunities of LIDO, the question arises as

to how manageable the mapping and ingestion process is for content providers who may have only recently started sharing their data in a wider service environment. To facilitate this process a mapping tool has been developed by the technical partner of the ATHENA project, the National Technical University of Athens.

Any kind of data provided in an XML format can be loaded into the system. The tool then visualizes, on the left, the incoming source data structure and, on the right, the LIDO target schema. The content provider can then map its source data fields through drag and drop to the target fields, including mapping of structural elements holding no data, and conditions for the mapping and concatenation of data values and constants. A helpdesk mailing list allows users to ask questions about the format and the tool, and to help each other. Combining a comprehensive metadata format with a customized technical solution for practical mapping is an exciting effort. It enables semantic interoperability of content from many different collections and from different management systems with different data structures. It is difficult to evaluate how the process will evolve over the next few months of the ATHENA project's activities and beyond, but some preliminary statements may be given here for discussion, both, positive and instructive. The overall mapping results are good and the questions on the helpdesk list comprehensive, so users appear to have grasped, from the material and the tool provided, both the LIDO schema and how to map to it.

Yet to get to a full and meaningful mapping that best reflects the source information in the target schema, several feedback loops are often needed between the local expert, who knows the source schema and content very well, and a LIDO expert who knows the LIDO structure in depth. This loop is considerably shortened by the ATHENA mapping tool, the result of a close cooperation between LIDO schema developers and technical imple-



Figure 2. ATHENA mapping tool

menters, which reflects the target schema very clearly. The process is considerably easier if the source schema is based on a documentation standard such as SPECTRUM or national standard. Moreover, features supporting data analysis and data value statistics, such as provided in the mapping tool, help immensely in this process.

## Conclusion

Overall it seems that it is both appropriate and simpler for content providers to map their data to a well-structured metadata format, instead of randomly choosing some corresponding field in a flat structure such as ESE.

Presently, LIDO serves in ATHENA as an intermediate layer between source formats and the Dublin Core-based ESE format. It thereby provides a more standardized representation of museum collections in Europeana. Since the ESE format does not support the fine granularity of museum information and fails to make a clear distinction between the museum object itself and its digital surrogate in an online service, standardized presentation helps to improve search and display quality considerably.

It will be crucial to see now the practical implementation of the new Europeana Data Model, EDM. EDM will supplement and enhance the currently used ESE model with a meta-structure that truly allows the LIDO format to be retrieved. It is a clear expectation that the implementation of this data model will significantly improve resource discovery, providing more precise search results that carry meaningful links to associated resources.

LIDO effectively prepares the ground for such new, data quality focused approaches. Used in conjunction with increasing opportunities to participate in linked data environments – as they are aimed at in the forthcoming EU-funded Linked Heritage project, this will enable museums to recontextualize their collections in a meaningful way and hence improve understanding of the collections within the greater cultural heritage context.

For full reference of LIDO visit <http://www.lido-schema.org/>.

Several training material can be found at <http://www.athenaeurope.org/index.php?en/159/training>.

# Final Report

## Digital Libraries: Recommendations and Challenges for the Future

Marco Ricolfi – Lynne Brindley – Claudia Dillman – Tarja Koskinen-Olsson  
Toby Bainton – Anne Bergman-Tahon – Jean-François Debarnot – Myriam Diocaretz  
Olav Stokkmo

i2010: Digital Libraries, High Level Expert Group, Copyright Subgroup

*Con questa traduzione si conclude la serie dei documenti presentati su «DigItalia» per far conoscere il lavoro del Gruppo di esperti (High level expert group) che ha lavorato dal 2006 alla fine del 2009 analizzando specifiche tematiche nell'ambito dell'iniziativa i2010: Digital Libraries della Commissione Europea.*

*In particolare sono stati tradotti a partire dal 2007 (1/2007, 2/2007, 2/2008, 2/2009) gli studi e le proposte sui problemi di copyright da affrontare nei progetti di digitalizzazione. Questo documento conclusivo è costituito dalle Raccomandazioni che i sottogruppi sulle specifiche tematiche (copyright, partenariati pubblico-privato, informazione scientifica) rivolgono ai responsabili della Commissione europea ed ai professionisti delle biblioteche digitali per realizzare in condizioni ottimali progetti di digitalizzazione, favorire accordi tra partner pubblici e privati, migliorare lo stato dell'informazione scientifica ed affrontare con consapevolezza le sfide per il futuro.*

### Relazione finale Biblioteche digitali: raccomandazioni e sfide per il futuro\*

Traduzione di Sonia Beraha per Qualitalia S.R.L

### Indice

1. Sintesi
2. Diritti di proprietà intellettuale: problematiche sul diritto d'autore per le biblioteche digitali
  - 2.1 Risultati e raccomandazioni
    - 2.1.1 Opere orfane: meccanismi a livello di Stati membri
    - 2.1.2 Opere orfane: Protocollo d'intesa sulle linee guida per la conduzione di ricerche accurate sulle opere orfane
    - 2.1.3 Opere esaurite: Accordo tipo
    - 2.1.4 Opere orfane e opere esaurite: banche dati e centri di gestione dei diritti; il progetto ARROW
    - 2.1.5 Conservazione digitale: copie multiple, migrazione e dispositivi tecnici di protezione
    - 2.1.6 Conservazione del contenuto Web e Web harvesting
  - 2.2 Sfide per il futuro
    - 2.2.1 Condizioni di base per la digitalizzazione e l'accessibilità online
    - 2.2.2 Privatizzazione delle opere di pubblico dominio nell'ambiente digitale
3. Partenariati pubblico-privato per la digitalizzazione e l'accessibilità online del patrimonio culturale

\* Il testo in inglese, redatto a dicembre del 2009, è consultabile all'indirizzo Web [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/hleg/reports/hlg\\_final\\_report09.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/hlg_final_report09.pdf).

- 3.1 Risultati e raccomandazioni
- 3.2 Sfide per il futuro: come promuovere correttamente i partenariati pubblico-privato
- 4. Informazioni scientifiche: accesso e conservazione
  - 4.1 Risultati e raccomandazioni
    - 4.1.1 Principi sulle informazioni scientifiche concordati dai rappresentanti degli editori, dalla comunità scientifica e dalle biblioteche. Aree di disaccordo
    - 4.1.2 Accesso e conservazione dei dati digitali delle ricerche
    - 4.1.3 Esperimenti di accesso aperto: PE-ER (Publishing and the Ecology of the European Research)
  - 4.2 Sfide per il futuro
- 5. Europeaana, la biblioteca digitale europea

Allegato 1. Membri e altri esperti del Gruppo di esperti di alto livello sulle Biblioteche digitali

Allegato 2. Elenco delle relazioni informative e dei documenti relativi al lavoro del Gruppo di esperti di alto livello sulle Biblioteche digitali<sup>1</sup>

## 1. Sintesi

La presente relazione illustra le proposte elaborate dal Gruppo di esperti di alto livello in materia di biblioteche digitali, nel periodo 2006-2009, nello specifico:

- riassumendo i risultati e le raccomandazioni relativamente ai quali il Gruppo ha raggiunto un consenso;
- identificando le questioni e le sfide che devono ancora essere affrontate, tramite la raccolta di prove, la sperimentazione, il dibattito tra le parti interessate e le iniziative di natura politica. Tali aspetti dovranno essere inclusi nell'agenda delle biblioteche digitali per i prossimi anni.

Il Gruppo di esperti di alto livello è stato creato dalla Commissione Europea all'inizio del 2006, nell'ambito dell'iniziativa sulle biblioteche digitali. Il Gruppo ha avuto il compito di consigliare la Commissione sul modo migliore di affrontare le sfide legali e tecniche a livello europeo e di contribuire ad una visione strategica comune delle biblioteche digitali europee.

Il Gruppo, composto da 20 membri nominati personalmente per le loro competenze specifiche in materia di biblioteche digitali, includeva esperti provenienti da istituti deputati alla conservazione della memoria (biblioteche, archivi, musei), titolari di diritti, fornitori di contenuti, aziende operanti nel settore dell'ICT (Information and communication technology, motori di ricerca, fornitori di tecnologia), organismi scientifici e di ricerca, istituzioni del mondo accademico. Inoltre, di volta in volta, specifici professionisti sono stati invitati a prendere parte alle assemblee plenarie e alle riunioni dei sottogruppi affinché apportassero il loro contributo nell'ambito di questioni specifiche. L'Allegato 1 riporta un elenco dei membri del Gruppo e degli altri esperti che hanno partecipato. Il Gruppo era presieduto da Viviane Reding, Commissario europeo per le telecomunicazioni e i media.

Nel corso del periodo 2006-2009, il Gruppo di esperti di alto livello si è riunito sei volte in sessione plenaria, una o due volte all'anno. Sono stati istituiti tre sottogruppi che si sono riuniti più volte allo scopo di preparare i dibattiti e le delibere dell'assemblea plenaria:

- il Sottogruppo sui diritti di proprietà intellettuale (diritti d'autore);
- il Sottogruppo sui partenariati pubblico-privato;
- il Sottogruppo sull'informazione scientifica.

<sup>1</sup> [Gli Allegati 1 e 2 al presente *Rapporto* sono consultabili all'indirizzo Web del testo in inglese, [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/hleg/reports/hlg\\_final\\_report09.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/hlg_final_report09.pdf)].

Per ognuna di queste aree, il Gruppo ha elaborato relazioni informative (elencate nell'Allegato 2) che sono state ampiamente divulgate nel sito Web della Commissione europea e tramite altri mezzi (Gruppo di esperti degli Stati membri sulla digitalizzazione e la conservazione digitale)<sup>2</sup>. Oltre a fornire una serie di linee-guida, la collaborazione tra le figure d'alto livello delle parti interessate ha condotto a risultati concreti. Infatti, dal lavoro del Gruppo di esperti di alto livello sugli aspetti relativi ai diritti d'autore è scaturita la redazione di un *Protocollo di intesa sulle opere orfane* sottoscritto dai rappresentanti delle istituzioni culturali e dei titolari dei diritti. Inoltre, tale lavoro ha spianato la strada ad ulteriori politiche per l'azione a livello comunitario in materia di opere orfane, azioni che sono attualmente al vaglio. Ha anche fornito un utile contributo per la preparazione e l'attuazione della *Raccomandazione della commissione sulla digitalizzazione, l'accessibilità online e la conservazione digitale del materiale culturale*.

Dal lavoro sui partenariati pubblico-privato sono scaturite indicazioni sui metodi pragmatici per incoraggiare gli investimenti privati nel settore della cultura digitale, preservando, allo stesso tempo, la natura pubblica dei beni di inestimabile valore affidati alle istituzioni culturali. La questione del pubblico dominio nell'ambiente digitale è stata chiaramente identificata come un'area che merita attenzione e che, eventualmente, dovrebbe diventare oggetto di un'azione politica, alla luce del rischio che la digitalizzazione possa "bloccare" le opere di pubblico dominio.

Per quanto concerne l'informazione scientifica, il lavoro svolto ha comportato un intenso dialogo tra editori scientifici e ricercatori, finalizzato a migliorare l'accesso alle pubblicazioni e ai dati scientifici. Questo lavoro ha portato allo sviluppo di alcuni principi generali condivisi e ha contribuito al lancio di una prima serie di azioni attuative, incluso un esperimento (progetto PEER) finalizzato a studiare l'effetto del libero accesso alle pubblicazioni scientifiche e alla ricerca europea in senso lato.

<sup>2</sup> [Le relazioni sono state in gran parte tradotte e pubblicate nei precedenti numeri di «DigItalia». Cfr.: i2010: Digital Libraries, High Level Expert Group, Copyright Subgroup, *Report on Digital Preservation, Orphan Works, and Out-of-Print Works: Selected Implementation Issues = Relazione sulla conservazione digitale, le opere orfane e le opere esaurite: Aspetti concreti legati all'attuazione*, «DigItalia», 1, p. 145-156, [http://digitalia.sbn.it/upload/documenti/digitalia20071\\_RICOLFI.pdf](http://digitalia.sbn.it/upload/documenti/digitalia20071_RICOLFI.pdf); i2010: Digital Libraries, High Level Expert Group, Copyright Subgroup, *Model agreement for a licence on digitisation of out of print works = Modello di accordo per una licenza di digitalizzazione di opere esaurite*, «DigItalia», 2 (2007), 2, p. 131-139, [http://digitalia.sbn.it/upload/documenti/digitalia20072\\_HLEG.pdf](http://digitalia.sbn.it/upload/documenti/digitalia20072_HLEG.pdf); European Digital Libraries, *Appendix to the Joint Report: Sector Reports = Appendice al Rapporto congiunto: Rapporti di settore*, «DigItalia», 3 (2008), 2, p. 131-139, [http://digitalia.sbn.it/upload/documenti/Digitalia20082\\_documenti\\_01.pdf](http://digitalia.sbn.it/upload/documenti/Digitalia20082_documenti_01.pdf); European Digital Libraries, *Sector-Specific Guidelines on Due Diligence Criteria for Orphan Works. Join Report = Linee guida settoriali sui criteri di dovuta diligenza per le opere orfane: rapporto congiunto*, «DigItalia», 3 (2008), 2, p. 135-141, [http://digitalia.sbn.it/upload/documenti/Digitalia20082\\_documenti\\_02.pdf](http://digitalia.sbn.it/upload/documenti/Digitalia20082_documenti_02.pdf); European Digital Libraries, *Memorandum of Understanding on Diligent Search Guidelines for Orphan Works = Protocollo d'intesa sulle linee guida per la conduzione di ricerche accurate sulle opere orfane*, «DigItalia», 3 (2008), 2, p. 142-145, [http://digitalia.sbn.it/upload/documenti/Digitalia20082\\_documenti\\_03.pdf](http://digitalia.sbn.it/upload/documenti/Digitalia20082_documenti_03.pdf); i2010: Digital Libraries High Level Expert Group, Copyright Subgroup, *Final Report on Digital Preservation, Orphan Works, and Out-of-Print Works = Relazione finale sulla conservazione digitale, le opere orfane e le opere esaurite*, «DigItalia», 4 (2009), 2, p. 126-154, <http://digitalia.sbn.it/upload/documenti/Digit%202-2009.pdf?l=it>].

## 2. Diritti di proprietà intellettuale: problematiche sul diritto d'autore per le biblioteche digitali

### 2.1 Risultati e raccomandazioni

#### 2.1.1 Opere orfane: meccanismi a livello di Stati membri

Il Gruppo di esperti di alto livello raccomanda quanto segue:

per quanto concerne le opere i cui titolari non siano identificabili o rintracciabili (opere orfane), si incoraggiano gli Stati membri a mettere a punto, ove applicabile, un meccanismo che consenta lo sfruttamento commerciale e non commerciale di tali opere a fronte di termini e compensi concordati, fermo restando che prima dell'utilizzo sia stata condotta una ricerca diligente nel paese di origine mirante ad identificarne e/o rintracciarne i titolari dei diritti. I meccanismi in essere negli Stati membri devono soddisfare determinati criteri in merito al contesto e all'utilizzo delle opere, ossia:

- la soluzione dovrebbe essere applicabile a tutti i tipi di opere;
- l'utente in buona fede deve condurre una ricerca diligente nel paese di origine prima di utilizzare l'opera;
- le migliori pratiche o le linee-guida relative a specifiche categorie di opere possono essere definite dalle parti interessate a seconda dei settori di appartenenza, ma tali linee-guida non devono essere integrate nella legislazione. Si invitano gli Stati membri a riconoscere le soluzioni adottate da altri Stati membri ove rispondano ai criteri prescritti, allo scopo di ottenere l'effetto *cross-border* voluto dall'iniziativa sulle biblioteche digitali. Pertanto, il materiale che può essere legalmente utilizzato in uno Stato Membro dovrebbe poter essere legalmente utilizzato anche in un altro Stato Membro.

#### 2.1.2 Opere orfane: Protocollo d'intesa sulle linee guida per la conduzione di ricerche accurate sulle opere orfane

Il *Protocollo di intesa sulle opere orfane* può essere definito come un "risultato derivato" del Gruppo di esperti di alto livello sulle biblioteche digitali. Il *Protocollo* è una sorta di autoregolamento sottoscritto da ventisette organizzazioni che rappresentano i titolari di diritti e le istituzioni culturali europee, i quali hanno accettato di attenersi ad una serie di linee-guida di diligenza per quanto concerne la ricerca dei titolari dei diritti. Inoltre, il *Protocollo* prevede che un'opera possa essere considerata orfana solo se siano stati applicati i rispettivi criteri, ivi inclusa la documentazione relativa al processo di ricerca che non ha rintracciato i titolari dei diritti.

Il *Protocollo di intesa*, unitamente alla *Raccomandazione (2.2.1)* relativa alle misure da adottare a livello nazionale, rappresentano le due pietre miliari per la soluzione della complessa questione delle opere orfane, per la quale sarebbe comunque auspicabile l'adozione di una legislazione a livello comunitario. Le parti interessate non sono giunte ad un accordo relativamente ai requisiti dei criteri di dovuta diligenza per la digitalizzazione di massa.

#### 2.1.3 Opere esaurite: Accordo tipo

Il Gruppo di esperti di alto livello ha sviluppato un *Accordo tipo* per la digitalizzazione e l'accesso online ai libri esauriti protetti da diritti d'autore o, in alternativa, l'accesso agli utenti autorizzati esclusivamente nell'ambito di reti chiuse.

L'*Accordo tipo* è stato concepito per essere adattabile ai diversi regimi giuridici e ai diversi modelli di gestione dei diritti in essere negli Stati membri dell'Unione Europea. Prima di tutto, il Gruppo di esperti di alto livello ha sviluppato un *Accordo tipo* che consente alla biblioteca di digitalizzare e fornire l'accesso alle opere esaurite solo agli utenti autorizzati all'interno di reti chiuse. La prima versione

dell'Accordo tipo non pone restrizioni di natura territoriale, ma l'accessibilità non è disponibile su reti aperte.

Quindi, il Gruppo di esperti di alto livello ha concepito una nuova versione dell'Accordo tipo che consente alle biblioteche di offrire l'accesso online completo. Inoltre, si ritiene che una biblioteca che offre l'accesso online a libri esauriti consentirà anche l'accesso alle stesse categorie di opere ad utenti autorizzati in reti chiuse. Pertanto, il secondo Accordo tipo che è stato elaborato contempla entrambe le opzioni.

#### **2.1.4 Opere orfane e opere esaurite: banche dati e centri di gestione dei diritti; il progetto ARROW**

Il Gruppo di esperti di alto livello raccomanda la creazione di Banche dati e di centri di gestione dei diritti per le opere orfane e per le opere esaurite. Il Gruppo di esperti di alto livello ha sviluppato una serie di *Principi cardine per le banche dati e i centri di gestione dei diritti delle opere orfane e anche dei Principi cardine per le banche dati e i centri di gestione dei diritti relativi alle opere esaurite*. Il Gruppo di esperti di alto livello raccomanda l'applicazione di tali principi, quando queste banche dati e questi centri di gestione dei diritti vengono creati.

Il progetto ARROW (Accessible registries of rights information and orphan works towards europea), nato nel contesto del programma eContentplus, non è altro che un primo passo in questa direzione. Il progetto è volto a fornire alle parti interessate le informazioni necessarie per chiarire la situazione dei diritti delle opere orfane e delle opere esaurite, al fine di consentirne la digitalizzazione e l'accessibilità da parte del pubblico. ARROW potrebbe costituire la base per ulteriori iniziative in quest'area.

#### **2.1.5 Conservazione digitale: copie multiple, migrazione e dispositivi tecnici di protezione**

Laddove uno Stato Membro abbia previsto

un'eccezione per consentire l'esecuzione di copie digitali di un'opera e ove le copie siano eseguite a scopo conservativo:

- alcune istituzioni, ovvero le biblioteche accessibili al pubblico, gli istituti di istruzione, i musei e gli archivi dovrebbero essere autorizzati a fare più di una copia (un numero illimitato di copie), nel caso in cui ciò fosse necessario per garantire la conservazione dell'opera. Successivamente, l'esecuzione di altre copie potrebbe essere autorizzata se e quando nuovi sviluppi tecnologici rendessero necessaria l'attuazione di tali misure, ma solo a scopo conservativo, ferma restando la tutela dell'identità e dell'integrità della singola pubblicazione;
- occorre coordinare le diverse iniziative di protezione attuate a livello regionale, nazionale e dell'Unione Europea, in modo da evitare duplicazioni tra le diverse iniziative, includendo le biblioteche nazionali incaricate del "deposito legale";
- nel caso delle biblioteche incaricate del deposito legale e per quanto concerne le opere nate in digitale dotate di un dispositivo di protezione, editori e biblioteche nazionali hanno concordato che tale dispositivo debba essere disabilitato nella copia del deposito legale (ad esempio per i fini delle biblioteche nazionali, ma non per l'accesso agli utenti finali), in modo da consentire, nel tempo, un accesso permanente e senza limitazioni al documento.

#### **2.1.6 Conservazione del contenuto Web e Web harvesting**

Il Gruppo di esperti di alto livello raccomanda che gli Stati membri affrontino in maniera prioritaria la questione del Web harvesting in quanto un quantitativo sempre crescente di materiale viene creato e aggiornato esclusivamente in formato elettronico come contenuto Web. L'obbligo di preservare il contenuto Web da parte di organizzazioni appositamente in-

caricate deve essere trattato come una questione inerente la legislazione del deposito legale. L'attuale legislazione comunitaria sui diritti di proprietà intellettuale, in particolare la Direttiva 2001/29 e gli articoli 9 e 5.2 (c) della stessa, tiene conto della conservazione del contenuto Web.

Gli aspetti seguenti relativi alla conservazione dei contenuti Web devono essere contemplati da tutte le legislazioni che disciplinano i diritti d'autore e i diritti correlati:

- a. laddove la legislazione sui diritti d'autore utilizza il termine "diritto di riproduzione", la conservazione a lungo termine del contenuto Web richiede azioni come la copia e la migrazione del materiale e questo deve essere chiarito nella terminologia;
- b. il contenuto Web potrebbe essere stato raccolto e depositato, nel paese di origine o all'estero, prima che nascesse l'obbligo di deposito legale e, allo scopo di consentire l'utilizzo di tale materiale depositato dal punto di vista del diritto d'autore, occorre trovare una soluzione alla questione della retroattività, con allo scopo di garantire il più possibile la conservazione del patrimonio culturale basato sul Web.

## 2.2 Sfide per il futuro

### 2.2.1 Condizioni di base per la digitalizzazione e l'accessibilità online

Come pubblicare online un maggior numero di opere coperte da diritto d'autore, soprattutto per quanto concerne le opere esaurite e le opere orfane, resta tuttora una delle sfide principali per le biblioteche digitali. Molto sarebbe stato ottenuto se le raccomandazioni del Gruppo di esperti di alto livello fossero state messe in pratica. Il Gruppo di esperti di alto livello ha lasciato aperta la questione di come riconsiderare il quadro legale della digitalizzazione del patrimonio culturale alla luce di tali sfide, soprattutto per bilanciare il principio dell'accesso alle informazioni con il riconoscimento dei diritti dei creatori, incluso il

compenso per il loro lavoro e la loro creatività. Anche in seno al Gruppo di esperti di alto livello, i titolari dei diritti insistono sul fatto che la digitalizzazione e l'accessibilità online debbano essere eseguite nel pieno rispetto delle normative che disciplinano i diritti d'autore, mentre le istituzioni culturali ritengono che sia necessario apportare riforme alla legislazione che disciplina i diritti d'autore e procedere ad un'ulteriore armonizzazione a livello europeo, se si vogliono creare le condizioni idonee ad una digitalizzazione su grande scala. È chiaro che si dovranno condurre ulteriori dibattiti su questo aspetto per comprendere meglio i diversi punti di vista.

### 2.2.2 Privatizzazione delle opere di pubblico dominio nell'ambiente digitale

Alcune istituzioni culturali o aziende private che effettuano la digitalizzazione dei contenuti culturali pretendono i diritti sulle copie digitalizzate di opere di pubblico dominio. Si tratta di stabilire se la digitalizzazione generi per se stessa nuovi diritti. Il livello di originalità necessario per la creazione di un diritto d'autore non è, allo stato attuale, armonizzato a livello europeo.

Il principio generale da applicare è che i lavori di pubblico dominio dovrebbero rimanere di pubblico dominio anche nell'ambiente digitale (cfr. paragrafo 3.1.6). In particolare, il Gruppo di esperti di alto livello raccomanda di non bloccare il materiale di pubblico dominio che sia stato digitalizzato dalle istituzioni con fondi pubblici, in quanto tale materiale deve continuare a svolgere il suo ruolo essenziale di fonte di creatività e innovazione.

## 3. Partenariati pubblico-privato per la digitalizzazione e l'accessibilità online del patrimonio culturale

### 3.1 Risultati e raccomandazioni

I partenariati pubblico-privato (PPP) potrebbero svolgere un ruolo importante nello sviluppo delle biblioteche europee (ivi inclusa

Europeana) e nell'attuazione di una strategia per la digitalizzazione, l'accessibilità online e la conservazione digitale della memoria collettiva europea. Se da un lato le biblioteche, gli archivi, i musei e le gallerie hanno già tutelato questa memoria collettiva e hanno maturato una certa esperienza in materia di identificazione delle risorse e delle esigenze degli utenti, i partner privati potrebbero contribuire con un valido apporto in termini di finanziamenti, tecnologia, software e di competenze necessarie per procedere ad una digitalizzazione su larga scala. La sinergia tra le parti consentirebbe di migliorare l'accesso del pubblico.

La maggior parte dei partenariati presi in esame come casi di studio si trovavano ancora nelle fasi iniziali di sviluppo e quindi era troppo presto per trarre conclusioni definitive in merito agli elementi di successo. A tal fine, occorre continuare ad osservare lo sviluppo dei partenariati analizzati.

Se da un lato vi è un ampio consenso sulla potenziale utilità dei PPP e sulla necessità di promuovere gli stessi, dall'altro, emergono chiaramente le difficoltà, i conflitti e i fallimenti associati ai PPP.

Il Gruppo di esperti di alto livello, sulla base di una serie di casi di studio, ha sviluppato un insieme di linee-guida e ha identificato alcune tematiche, allo scopo di consentire agli attori pubblici e privati di trarre le proprie conclusioni in merito all'aiuto che la forma di partenariato pubblico-privato potrebbe comportare nei loro casi specifici.

Il Gruppo di esperti di alto livello raccomanda quanto segue:

- 3.1.1. I partner devono definire chiaramente i loro obiettivi strategici e i vantaggi che i cittadini potrebbero trarre dal progetto.
- 3.1.2. Il partenariato dovrebbe sfruttare al massimo l'esperienza e la competenza dei partner, apportando contributi complementari.

3.1.3. Il partenariato dovrebbe massimizzare l'accesso da parte del pubblico e prendere in considerazione gli aspetti legati alla conservazione e alla sostenibilità a lungo termine.

3.1.4. Ambiente finanziario: la legislazione volta a supportare il finanziamento del patrimonio culturale attraverso agevolazioni fiscali ai partner privati dovrebbe essere applicata in maniera più capillare ai progetti di digitalizzazione.

3.1.5. Diritti di proprietà intellettuale: i partenariati devono essere integrati nel contesto delle leggi che disciplinano i diritti d'autore.

3.1.6. Pubblico dominio: il contenuto analogico di pubblico dominio dovrebbe rimanere di pubblico dominio anche nell'ambiente digitale. Ove sia necessario applicare restrizioni di accesso e di utilizzo per l'utente al fine di rendere disponibile il contenuto digitale, tali restrizioni dovranno essere applicabili solo per un periodo di tempo limitato.

3.1.7. Esclusività: sono da evitare gli accordi a carattere esclusivo per la digitalizzazione e la distribuzione dei beni digitali delle istituzioni culturali. Laddove si rendesse necessario sottoscrivere accordi esclusivi per fornire un servizio di pubblico interesse, tali accordi di esclusiva dovranno essere limitati nel tempo, trasparenti e soggetti regolarmente a revisione.

3.1.8. Riutilizzo: le istituzioni culturali dovrebbero fare il possibile per rispettare i principi della Direttiva Europea 2003/98/CE in materia di riutilizzo dell'informazione pubblica.

3.1.9. *Governance*: i PPP, oltre ad essere trasparenti e attendibili, dovranno includere accordi di *governance* nell'ambito di un contratto formale sottoscritto dalle parti.

### 3.2 Sfide per il futuro: come promuovere correttamente i partenariati pubblico-privato

In Europa, i PPP non hanno preso piede come metodo comune e diffuso di digitalizzazione dei contenuti. Da una parte, è chiaro che i PPP sono essenziali per garantire la digitalizzazione e l'accessibilità online al patrimonio culturale. Per questo motivo, devono essere promossi e incoraggiati e l'investimento dei partner privati deve essere adeguatamente ricompensato. Dall'altra parte, occorre riuscire in questo intento rispettando pienamente, allo stesso tempo, i titolari dei diritti e, allo stesso tempo, garantendo il valore dei beni pubblici conservati dalle istruzioni culturali e potenziando l'accesso ai contenuti.

## 4. Informazione scientifica: accesso e conservazione

### 4.1 Risultati e raccomandazioni

#### 4.1.1 Principi sull'informazione scientifica concordati dai rappresentanti degli editori, la comunità scientifica e le biblioteche.

##### Area di disaccordo

Il Gruppo di esperti di alto livello ha operato come un forum dove i ricercatori, gli enti finanziatori e gli editori scientifici si sono incontrati per esprimere i loro punti di vista con lo scopo di identificare soluzioni pratiche e condivise volte a migliorare l'accesso alle informazioni scientifiche. Il metodo adottato prevedeva innanzitutto l'identificazione delle aree di accordo e di disaccordo tra le parti. Le aree di accordo si riferiscono principalmente ai principi generali (e cioè il potenziamento della divulgazione e dell'accesso, la libertà dei ricercatori, un compenso equo per la pubblicazione), ai dati delle ricerche e alla conservazione. Uno dei principali punti di disaccordo, invece, concerne l'opzione del deposito obbligatorio degli articoli scientifici in archivi Open Access dopo un periodo di embargo. Un altro aspetto che ha registrato una certa divergenza di opinioni è la disponibilità delle pubblicazioni scientifiche, soprattutto in relazione al

costo di accesso alle pubblicazioni periodiche pregresse (le biblioteche optano per un prezzo che si abbassa con il passare del tempo).

#### 4.1.2 Accesso e conservazione dei dati digitali delle ricerche

- I depositi digitali nei quali sono raccolti e gestiti i contenuti scientifici comprendono pubblicazioni ma, in misura crescente, anche dati primari e dati elaborati. Infatti, stanno aumentando i collegamenti tra le pubblicazioni e i dati. La possibilità di utilizzare e riutilizzare tali dati il più liberamente possibile è fondamentale ai fini dell'innovazione e del progresso scientifico.
- I principali attori del mondo della scienza, della scienza dell'informazione e della rappresentanza politica a livello nazionale e comunitario devono mettere a punto un generale quadro di riferimento che includa modelli di conservazione e di *funding/business* sostenibili. Lo scopo è quello di definire i ruoli e le responsabilità nella costruzione di un'infrastruttura informativa digitale europea che consenta l'accesso e il riutilizzo dei dati delle ricerche, ed assicuri anche la conservazione a lungo termine.
- La costruzione di un'infrastruttura adeguata richiede ingenti risorse in termini di impegno e di denaro. Considerando la crescente importanza della conservazione dei dati delle ricerche, le organizzazioni che finanziano le ricerche a livello nazionale ed europeo dovrebbero considerare di assegnare una parte più consistente dei loro budget alla conservazione dei dati delle ricerche.
- La Commissione Europea dovrebbe sviluppare ulteriormente le iniziative esistenti.

#### 4.1.3 Esperimenti di Open Access: PEER (Publishing and the ecology of the european research)

Dalle discussioni in seno al Gruppo di esperti di alto livello è scaturito un esperimento condiviso in materia di test pratici volti a racco-

gliere prove dell'impatto dell'Open Access ritardato (il cosiddetto "periodo di embargo"). Il progetto PEER, finanziato all'interno del programma eContentplus, è una collaborazione attiva tra editori, depositi digitali e ricercatori. Ha lo scopo di monitorare gli effetti del deposito sistematico, su larga scala, dei manoscritti degli autori già sottoposti a *peer-review* (Green Open Access), sull'accessibilità per il lettore, sulla visibilità dell'autore e sulla vitalità della pubblicazione, nonché sulla più ampia ecologia della ricerca e della pubblicazione in Europa. Gli editori partecipanti contribuiscono al progetto con circa 300 pubblicazioni. Il progetto è attualmente in corso, ma si dovrà attendere il 2011 per vederne i risultati.

#### 4.2 Sfide per il futuro

Il Gruppo di esperti di alto livello ha definito per gli anni a venire l'ordine del giorno di seguito descritto, che comprende una serie di argomenti da discutere e di raccomandazioni in materia di informazioni scientifiche in Europa:

Nuovi paradigmi dell'informazione scientifica:

- il percorso verso un accesso aperto alle pubblicazioni scientifiche;
- le barriere legali e pratiche che ostacolano le iniziative di estrazione dei dati e i mezzi per superare tali barriere;
- gli esperimenti devono essere condotti utilizzando metodi diversi di revisione ed estrazione dei dati su tutto il testo degli articoli scientifici.

Condivisione dei dati:

- compenso adeguato per i ricercatori che condividono i propri dati;
- espletamento dei mandati di ricerca dei dati da parte delle organizzazioni di finanziamento della ricerca;
- gli esperimenti devono essere condotti con metodi alternativi di revisione e con l'estensione dei dati dell'intero testo degli articoli scientifici.

Conservazione digitale:

- deposito legale dell'informazione scientifica digitale nell'Unione Europea;
- responsabilità delle biblioteche nazionali e di ricerca in materia di conservazione dell'informazione scientifica nell'ambito del loro ruolo e dei loro compiti nell'era digitale, correlata alle esigenze di finanziamento;
- meccanismi per il finanziamento della conservazione digitale delle ricerche attuati dalle organizzazioni di finanziamento (supporto dell'infrastruttura di conservazione e di interventi una tantum);
- gli esperimenti devono essere condotti a livello comunitario, allo scopo di assicurare la qualità e l'affidabilità dei depositi e sviluppare una struttura per la revisione e la certificazione di tali depositi.

## 5. Europea, la biblioteca digitale europea

L'emergere e il proliferare delle biblioteche digitali è una caratteristica tipica del nuovo millennio. Le biblioteche digitali sono create dalle istituzioni culturali "tradizionali" e dai fornitori di contenuti che sviluppano le loro attività digitali, oppure da nuove organizzazioni che prima non operavano con i contenuti in formato analogico. Le dimensioni, le tipologie e le finalità di tali organizzazioni e delle biblioteche digitali sono anch'esse alquanto diverse: dimensioni globali, internazionali, nazionali o locali; strutture pubbliche, private o miste; finalità no profit di accessibilità online rispetto ad obiettivi essenzialmente commerciali. Nell'ambito di tale processo, le tecnologie digitali hanno una capacità senza precedenti di rimodellare le relazioni tra le organizzazioni stesse. I confini geografici sono superati, è possibile l'apertura e la connettività. Le differenze e i confini tra i diversi tipi di istituzioni culturali (per esempio tra le biblioteche, gli archivi e i musei) tendono a sparire. Il mercato svolge un ruolo essenziale nella modifica dell'economia della cultura digitale e le relazioni tra enti pubblici e imprese private assumono nuove forme.

Una visione a lungo termine della biblioteca digitale globale può essere utile per guidare questo processo: una rete internet che fornisca a tutti l'accesso alla conoscenza universale per contribuire all'istruzione, alla ricerca, all'apprendimento costante, alla creatività, all'innovazione e allo sviluppo sociale ed economico. Vi è la necessità di una politica pubblica che accompagni lo sviluppo delle biblioteche digitali e vi è anche la necessità di una biblioteca digitale posizionata a livello europeo:

Europeana. Europeana rappresenta il risultato più visibile dell'iniziativa sulle Biblioteche digitali e dovrebbe essere ulteriormente consolidata e sviluppata. Le istituzioni della UE, gli Stati membri e le parti in gioco dovrebbero tutti fornire pieno sostegno ad Europeana, in qualità di "campione europeo" nell'arena globale della cultura digitale. In questo modo, Europeana potrebbe trarre beneficio dall'utilizzo e dalla promozione delle raccomandazioni del Gruppo di esperti di alto livello.





# Conferenza “Meeting with Manuscripts, today and tomorrow”

Roma, 26–28 maggio 2010

**Valentina Longo**

*Biblioteca nazionale centrale di Roma*

La Conferenza, organizzata dal Manuscript Librarians Expert Group di LIBER (Ligue des Bibliothèques Européenne de Recherche) è stata ospitata dalla Biblioteca nazionale centrale di Roma. Il Gruppo, formatosi a Stoccolma nel 2000 con lo scopo di fornire ai bibliotecari conservatori di manoscritti uno spazio dedicato dove scambiarsi informazioni ed esperienze relative alle problematiche della conservazione, tutela e valorizzazione del patrimonio manoscritto ad essi affidato, era ormai al suo quarto appuntamento, che ha visto riuniti più di 80 colleghi ed esperti del settore provenienti da tutta Europa. Le numerose relazioni, distribuite in tre intense giornate dedicate anche a incontri, dibattiti e visite guidate a biblioteche ed archivi romani, si sono focalizzate sulla presentazione e discussione di esperienze e progetti sviluppati sia a livello locale sia a partecipazione internazionale, con particolare attenzione alle prospettive future, strettamente connesse con i progressi della tecnologia.

I principali temi affrontati durante la conferenza sono infatti legati allo sviluppo e alla diffusione delle tecnologie digitali, al loro impatto nel campo dell'informazione e della conoscenza in generale e della cultura manoscritta in particolare, e di conseguenza al modo in cui d'ora in avanti dovrà essere riconsiderata la figura professionale del curatore nel suo rapporto con i manoscritti e con i loro utenti.

Dopo il saluto del Direttore della Biblioteca

nazionale Osvaldo Avallone, che ha auspicato una fattiva collaborazione tra le istituzioni per la valorizzazione del patrimonio manoscritto, oggi più facile attraverso l'uso delle nuove tecnologie ma sempre penalizzata dalla carenza di risorse, e dopo la presentazione delle collezioni manoscritte della Biblioteca da parte di Francesca Niutta, i lavori della conferenza sono entrati nel vivo con la prima sessione dedicata alle presentazioni dei *National Reports*: sono state illustrate le collezioni di manoscritti e le recenti iniziative nel campo della catalogazione, digitalizzazione e pubblicazione online da parte dei colleghi delle biblioteche nazionali di Belgio, Bulgaria, Repubblica Ceca, Danimarca, Lituania, Olanda, Norvegia, Portogallo, Svizzera e della Biblioteca Apostolica Vaticana. I *Reports*, unitamente a quelli presentati nelle precedenti conferenze e ai relativi aggiornamenti, sono consultabili sul sito web del Gruppo (<http://liber-manuscripts.kb.nl/>).

L'aspetto più innovativo della conferenza è emerso dalle relazioni presentate nella seconda sessione, dedicata ai *born digital materials*: si tratta di una definizione che riunisce l'eterogeneo materiale digitale di cui è composto l'archivio personale dello studioso moderno – non più o non soltanto documenti cartacei, ma soprattutto testi elettronici, immagini digitali, documenti sonori – e che inizia a costituire una cospicua parte delle recenti acquisizioni delle biblioteche. Il fatto che questo tipo di materiale venga oggi annoverato tra le col-

lezioni di manoscritti non sembra aver suscitato alcuna perplessità tra gli addetti ai lavori, che considerano i testi nati digitali e in molti casi mai riprodotti su carta come i manoscritti del futuro, ponendo l'accento sull'unicità delle opere piuttosto che sul supporto su cui esse sono registrate.

Tutti i relatori intervenuti in questa sessione hanno cercato di individuare le migliori pratiche per gestire e preservare per le future generazioni i documenti nati digitali sia di argomento letterario sia storico, e la necessità di avvalersi dei nuovi scenari tecnologici allo scopo di ottimizzare l'accessibilità e ampliare il pubblico delle fonti manoscritte disponibili online. Nel caso di *born digital material* il curatore si trova ad affrontare una varietà di questioni, che vanno dalla gestione del materiale, conservato su supporti diversi da salvaguardare e verificare periodicamente, alla catalogazione e alla predisposizione di strumenti che ne garantiscano la consultabilità da parte dell'utenza sia a livello locale sia in remoto, tenendo tuttavia sempre presenti le questioni di natura legale (diritti d'autore e di sfruttamento da parte dell'Ente conservatore, tutela della privacy, garanzia di autenticità del documento). Tutto ciò comporta un necessario adeguamento tanto della formazione professionale del bibliotecario conservatore quanto della struttura che ospita tali raccolte, sotto l'aspetto sia della conservazione (copie master, server dedicati e predisposti alla condivisione di immagini sul web), sia della consultazione (postazioni multimediali adeguate in sede, realizzazione di archivi consultabili online). Tale adeguamento richiederebbe investimenti consistenti, piuttosto improbabili in un'epoca di tagli ai fondi per la cultura, almeno per quanto riguarda il nostro Paese. Nella relazione introduttiva di Francesca Niutta è stato ad esempio ricordato il caso del progetto Biblioteca Nonantolana Virtuale – teso a ricostruire virtualmente la biblioteca medievale dell'Abbazia di Nonantola, di cui fanno parte 45 manoscritti oggi conservati

presso la Biblioteca nazionale di Roma e circa altrettanti dispersi in numerose biblioteche italiane ed estere –, presentato alla conferenza LIBER del 2007 ma ancora in attesa di completamento per mancanza di adeguati finanziamenti. Non a caso i progetti FutureArch e BEAM, illustrati da Susan E. Thomas della Bodleian Library di Oxford, sono consistentemente finanziati da una fondazione privata, mentre altri istituti di conservazione confidano nella collaborazione degli stessi produttori di questi archivi "ibridi", come recita l'eloquente titolo della relazione esposta da Claus Jensen della Biblioteca Reale danese: *A pragmatic solution to the acquisition and archiving of born-digital material: letting scholars auto-archive*.

Nella terza sessione, dal titolo *Networks of medieval and modern manuscripts within and outside their context: new cooperation models for librarians, archivists and scholars*, sono state illustrate alcune reti di manoscritti digitalizzati progettate in collaborazione tra istituzioni diverse. Il progetto Europeana Regia (che mira a ricostruire virtualmente le biblioteche delle principali abbazie carolingie dei secoli VIII-IX, la biblioteca di Carlo V e quella dei Re d'Aragona) e quello dei manoscritti del *Roman de la Rose* sono stati presentati da Thierry Delcourt della Bibliothèque nationale de France, mentre la raccolta della corrispondenza e dei diari di viaggio dell'esploratore e naturalista Alexander von Humboldt è stata illustrata da Toni Bernhart della Universität der Künste zu Berlin. Anche nel caso di questi progetti si sono dimostrati fondamentali tanto lo sviluppo di modelli di cooperazione tra diversi istituti di conservazione quanto il reperimento dei necessari finanziamenti: Europeana Regia si avvale in parte di finanziamenti UE e in parte di fondi stanziati dai singoli Paesi partner.

L'ultima sessione ha riguardato l'impatto che le nuove tecnologie hanno avuto sulla figura professionale del bibliotecario conservatore di manoscritti e sul pubblico che utilizza questo materiale.

Il titolo *Old and new audiences for manuscripts* rimanda all'esigenza avvertita da molti colleghi di favorire la nascita di nuove categorie di pubblico grazie alla realizzazione di percorsi multimediali che possano avvicinare ai manoscritti un'utenza non specialistica ma potenzialmente interessata. La realizzazione di strumenti – trascrizioni, descrizioni catalografiche, approfondimenti di carattere storico, artistico o letterario, collegamenti ipertestuali – che facilitino e rendano più stimolante la consultazione delle collezioni di manoscritti digitalizzati non deve tuttavia far dimenticare che l'immagine digitale, per quanto fedele e dettagliata, non potrà in nessun caso sostituire completamente il codice manoscritto e che la digitalizzazione difficilmente potrà alleggerire gli impegni più pressanti del curatore di manoscritti, legati alla conservazione e allo studio degli originali. Com'è risultato evidente dalle relazioni esposte, l'aggiornamento professionale va oggi considerato anche alla luce delle nuove sfide aperte dallo sviluppo dell'Information Technology, senza tuttavia tralasciare per questo le discipline cosiddette tradizionali, che rimangono imprescindibili per la formazione del bibliotecario curatore di manoscritti.

In occasione della Conferenza si è anche riunito lo Steering Committee del Gruppo, alla

cui presidenza Jutta Weber (Staatsbibliothek zu Berlin) è succeduta ad André Bouwman (Universiteitsbibliotheek Leiden), mentre il nuovo segretario è Ivan Boserup (Kongelige Bibliothek København); in sostituzione di Francesca Niutta, membro del comitato fin dalla sua costituzione, è subentrata Anna Gudayol (Biblioteca de Catalunya). In seguito alla riorganizzazione della struttura di LIBER, dall'estate 2009 il Gruppo fa parte del Working Group on Heritage Collections (WGHC), e la nuova presidente nella relazione conclusiva ha annunciato l'intenzione di stabilire nuovi e proficui rapporti con il CERL (Consortium of European Research Libraries) allo scopo di offrire una maggiore visibilità alle collezioni digitali di manoscritti antichi e moderni. Il LIBER Manuscript Librarians Expert Group, attraverso il sito internet e le liste di discussione che vi saranno ospitate, intende proporsi come un forum a livello europeo dove non solo gli addetti ai lavori ma anche tutti gli interessati potranno affrontare questioni relative alla conservazione e valorizzazione delle raccolte manoscritte. A conclusione dei lavori è stato annunciato che le conferenze del Gruppo avranno d'ora in poi una cadenza biennale, e che pertanto la prossima avrà luogo nella primavera del 2012 e sarà ospitata a Parigi dalla Bibliothèque Nationale.

# Conferenza “CLEF 2010: Conference on Multilingual and Multimodal Information Access Evaluation”

Padova, 20-23 Settembre 2010

**Maristella Agosti – Nicola Ferro**

*Università degli studi di Padova*

L'Università degli studi di Padova ha organizzato e ospitato dal 20 al 23 settembre 2010 la conferenza internazionale “CLEF 2010: Conference on Multilingual and Multimodal Information Access Evaluation”<sup>1</sup>. La conferenza si occupa della *valutazione sperimentale* dei sistemi di reperimento dell'informazione con una particolare attenzione ai problemi del multilinguismo, multimedialità e multi modalità. L'esempio più noto di sistema di reperimento dell'informazione sono i motori di ricerca nel Web ma tali sistemi sono anche usati come componenti di sistemi più ampi, quali i portali aziendali o le biblioteche digitali, con la finalità di fornire un accesso efficace ai contenuti gestiti. La valutazione sperimentale consiste nell'usare collezioni realistiche di documenti quali, ad esempio, pagine HTML prese dal Web, blog e Wikipedia, record catalografici collezionati da cataloghi in linea o biblioteche digitali, brevetti registrati a livello internazionale in varie lingue, oppure immagini mediche corredate dalle corrispondenti diagnosi, e sperimentare i sistemi sviluppati dai partecipanti all'iniziativa su tali documenti al fine di misurarne le prestazioni e comprendere quali siano gli algoritmi e le tecniche più efficaci.

Il CLEF (Cross-Language Evaluation Forum)<sup>2</sup> è un'importante iniziativa a livello internazionale che si è tenuta con regolarità ogni anno a partire dal 2000 e a cui partecipano i più importanti gruppi di ricerca a livello europeo ed internazionale, circa centocinquanta, così come aziende ed industrie leader nel settore. Assieme a TREC (Text REtrieval Conference)<sup>3</sup>, organizzato dal National Institute of Standards and Technology (NIST) negli Stati Uniti, e a NTCIR (NII Test Collection for IR Systems)<sup>4</sup>, organizzato dal National Institute of Informatics (NII) in Giappone, CLEF rappresenta una delle tre più importanti iniziative di valutazione sperimentale a livello mondiale. Nel corso degli anni, CLEF ha prodotto un miglioramento significativo e misurabile nelle prestazioni dei motori di ricerca multilingue, che sono quasi raddoppiate, offrendo agli utenti risultati di migliore qualità e riducendo la necessità di scorrere lunghe liste di risultati.

CLEF 2010 rappresenta un'innovazione radicale rispetto al formato classico di CLEF ed un esperimento per comprendere meglio come debbano essere strutturate le campagne di valutazione sperimentale di “prossima ge-

<sup>1</sup> CLEF 2010, <http://www.clef2010.org/>.

<sup>2</sup> CLEF, <http://www.clef-campaign.org/>.

<sup>3</sup> TREC, <http://trec.nist.gov/>.

<sup>4</sup> NTCIR, <http://research.nii.ac.jp/ntcir/>.

nerazione". Infatti, dopo dieci anni di successi, si è dovuto affrontare il problema di come innovare CLEF pur mantenendo il suo focus tradizionale incentrato sulla valutazione dei sistemi di reperimento dell'informazione. La scelta, dopo vivaci discussioni all'interno della comunità, è stata rendere CLEF un evento indipendente di quattro giorni non più organizzato congiuntamente alla ECDL (European Conference on Digital Libraries)<sup>5</sup> dove CLEF è sempre stato un workshop della durata di due giorni e mezzo. I primi due giorni sono stati una conferenza vera e propria con articoli selezionati in base ad un rigoroso processo di revisione, relatori invitati e tavole rotonde, il tutto incentrato sull'avanzamento metodologico e scientifico della valutazione sperimentale nel settore del reperimento dell'informazione. I secondi due giorni sono stati costituiti da laboratori e workshop organizzati in sessioni parallele che hanno riguardato la presentazione e discussione delle attività di valutazione condotte con sistemi reali e dei risultati ottenuti.

L'evento è stato organizzato con la collaborazione di diverse istituzioni a livello nazionale ed internazionale: la prof. Maristella Agosti, Università degli studi di Padova, e il prof. Maarten de Rijke, Università di Amsterdam, hanno svolto il ruolo di *general chair*; il dr. Nicola Ferro, Università degli studi di Padova, e il prof. Alan Smeaton, Università di Dublino, sono stati i *program chair*; la dr. Donna Harman del National Institute of Standards and Technology (NIST), USA, e il dr. Martin Braschler dell'Università di Scienze Applicate di Zurigo hanno svolto il ruolo di *labs co-chair*; infine il dr. Emanuele Pianta del Centre for the Evaluation of Language Communication

Technologies (CELCT) di Trento ha ricoperto il ruolo di *organization chair* mentre il dr. Khalid Choukri della Evaluations and Language resources Distribution Agency (ELDA) di Parigi ha svolto il ruolo di *resource chair*.

Circa 140 ricercatori provenienti da Europa, Stati Uniti, Asia e Sud America hanno partecipato all'evento, la maggior parte per la sua intera durata di quattro giorni, e circa una ventina tra studenti, dottorandi e giovani ricercatori hanno ricevuto una borsa di studio per prendere parte alla conferenza, grazie alle sponsorizzazioni ricevute. Le presentazioni, le foto e i filmati degli interventi tenuti durante l'evento sono disponibili sul sito Web di CLEF 2010.

### CLEF 2010: la conferenza

La conferenza mira all'avanzamento della ricerca sulla valutazione sperimentale dei sistemi informativi multilingue e multimediali al fine di supportare individui, organizzazioni e comunità che progettano, sviluppano, utilizzato e migliorano tali sistemi. Sono stati sollecitati articoli che esplorassero le esigenze e le pratiche correnti per l'accesso all'informazione multilingue e multimediale, che studiassero nuove metriche e metodologie di valutazione e che discutessero e proponessero nuove direzioni per le future attività di valutazione. In aggiunta, si è incoraggiata la sottomissione di articoli che analizzassero i risultati ottenuti in dieci anni di attività di CLEF conducendo esperimenti approfonditi sulle collezioni e i dati prodotti da CLEF nel corso del tempo. A questo fine i dati sperimentali degli ultimi dieci anni di CLEF sono stati resi disponibili in linea tramite il sistema DIRECT (Distributed Information Retrieval Evaluation Campaign

<sup>5</sup> ECDL è stato, per quattordici anni, l'evento scientifico di riferimento in Europa per la comunità delle biblioteche digitali. In occasione del suo quindicesimo anno, la conferenza è stata rinominata *Theory and Practice of Digital Libraries* (TPDL). La prossima edizione di TPDL si terrà a Berlino dal 25 al 29 settembre 2011.

TPDL 2011, <http://www.tpd2011.org/>

Tool)<sup>6</sup> e accessibili a seguito della registrazione nel sistema. Infine, è stato predisposto un ampio comitato di programma che coprisse non solo le aree tradizionalmente parte di CLEF ma che esplorasse anche nuove aree, come ad esempio la visualizzazione di grandi moli di dati.

Gli articoli accettati per la conferenza hanno riguardato attività di ricerca sulle risorse, gli strumenti e i metodi necessari per la valutazione sperimentale, così come la creazione di collezioni sperimentali realistiche e vicine alle necessità degli utenti.

In aggiunta, due relazioni invitate hanno illustrato prospettive future per la valutazione sperimentale sia dal punto di vista accademico sia dal punto di vista industriale. Il prof. Norbert Fuhr, Università di Duisburg-Essen in Germania, ha esplorato il tema *IR Between Science and Engineering, and the Role of Experimentation*, mentre il dr. Ricardo Baeza-Yates, Yahoo! Research e Universitat Pompeu Fabra in Spagna, ha illustrato le problematiche inerenti *Retrieval Evaluation in Practice*.

Il programma della conferenza ha compreso anche riepiloghi sulle attività condotte nei diversi "laboratori", spiegate poi nel dettaglio nei due giorni successivi, e interventi sui risultati delle altre campagne di valutazione a livello internazionale: TREC (Text REtrieval Conference)<sup>7</sup> negli Stati Uniti, NII-NACSIS Test Collection for IR Systems (NTCIR)<sup>8</sup> in Giappone, INitiative for the Evaluation of XML Retrieval (INEX)<sup>9</sup> in Australia, Russian Information Retrieval Evaluation Seminar (ROMIP)<sup>10</sup> in Russia, e Forum for Information Retrieval Evaluation (FIRE)<sup>11</sup> in India.

Al fine di stimolare ulteriormente la discussione e coinvolgere i partecipanti, sono state organizzate due tavole rotonde. La prima, *The Four Ladies of Experimental Evaluation*, ha visto coinvolte le quattro ricercatrici che hanno creato e coordinato le principali iniziative di valutazione degli ultimi venti anni: dr. Donna Harman (National Institute of Standards and Technology, NIST, USA), prof. Noriko Kando (National Institute of Informatics, NII, Japan), prof. Mounia Lalmas (University of Glasgow, UK) e dr. Carol Peters (Consiglio Nazionale delle Ricerche, ISTI-CNR, Pisa, Italia). In particolare, si è discusso di quali risultati siano stati effettivamente raggiunti fino ad oggi e quali ancora debbano essere conseguiti.

La seconda tavola rotonda, *A PROMISE for Experimental Evaluation*, è stata presentata dai membri del progetto PROMISE (Participative Research labORatory for Multimedia and Multilingual Information Systems Evaluation), una rete di eccellenza finanziata nel settimo programma quadro della Commissione Europea. PROMISE ambisce a far progredire la valutazione sperimentale dei sistemi di accesso all'informazione multilingue, multimediale e multi modale. L'obiettivo finale del progetto è fornire e rendere disponibili a livello internazionale un ambiente e un'infrastruttura software che unifichino i dati prodotti dalla valutazione sperimentale, la conoscenza da essi estratta, le metodologie di valutazione, gli strumenti per effettuare e automatizzare la valutazione e la comunità di utenti ad essa interessata, accrescendo l'utilizzo di tali dati e strumenti.

I risultati e gli interventi dei due giorni di con-

<sup>6</sup> DIRECT, <http://direct.dei.unipd.it/>. Maristella Agosti – Giorgio Maria Di Nunzio – Marco Dussin – Nicola Ferro, *10 Years of CLEF Data in DIRECT: Where We Are and Where We Can Go*, in: *Proceedings 3rd International Workshop on Evaluating Information Access (EVIA 2010)*, edited by Tetsuya Sakay, Mark Sanderson, William Webber, Tokyo: National Institute of Informatics, 2010, p. 16-24.

<sup>7</sup> Vedi nota 3.

<sup>8</sup> Vedi nota 4.

<sup>9</sup> INEX, <http://www.inex.otago.ac.nz/>.

<sup>10</sup> ROMIP, <http://romip.ru/en/>.

<sup>11</sup> FIRE, <http://www.isical.ac.in/~fire/>.

ferenza sono stati pubblicati negli atti editi da Springer<sup>12</sup>.

## CLEF 2010: i laboratori e i workshop

I laboratori continuano ed espandono l'approccio tradizionale di CLEF. Sono stati previsti due tipi diversi di laboratori: attività di valutazione ed analisi comparativa delle prestazioni e workshops che esplorino nuove problematiche per la valutazione sperimentale. È stato formato un comitato di selezione dei laboratori per valutare le proposte di laboratorio e scegliere quelle più innovative per CLEF 2010. Sono stati adottati diversi criteri di selezione tra cui: appropriatezza della metodologia adottata, fattibilità dell'attività proposta, riferimento ad esigenze di utenti ben definite, esistenza di un caso d'uso di interesse, numero di potenziali partecipanti, prospettiva di evoluzione della proposta, dimensione e scala degli esperimenti, inter-disciplinarietà, originalità e possibilità di riutilizzare i risultati. I risultati delle attività dei laboratori sono riportate in una pubblicazione separata, le *working notes*, distribuita durante la conferenza e disponibile in linea<sup>13</sup>. È auspicato che i laboratori diano anche origine a pubblicazioni successive all'evento, come ad esempio articoli su rivista.

Sono state offerte le seguenti attività di valutazione e workshop:

- CLEF-IP: un'attività di valutazione riguardante la proprietà intellettuale e i brevetti multilingue;
- ImageCLEF: un'attività di valutazione riguardante la ricerca e recupero di immagini e testo, con particolare riferimento ad

immagini di tipo medico, fotografie prese da Flickr e immagini di Wikipedia;

- PAN: un'attività di valutazione sull'identificazione del plagio e dell'abuso dei software sociali, con particolare riferimento a Wikipedia;
- RespubliQA: un'attività di valutazione focalizzata sul *question answering* utilizzando una collezione multilingua di documenti del parlamento europeo;
- WePS: un'attività di valutazione focalizzata sulla ricerca di persone nel Web e sulla gestione della propria reputazione su Web;
- CriES: un workshop finalizzato ad esplorare la valutazione della ricerca di competenze e conoscenze nelle reti sociali;
- LogCLEF: un workshop finalizzato all'analisi e classificazione delle interrogazioni al fine di comprendere il comportamento dell'utente in un contesto multilingua, utilizzando i log di TEL (The European Library)<sup>14</sup> e del motore di ricerca DBS (Deutscher Bildungsserver).

## CLEF 2010: l'evento

I primi due giorni, la conferenza, sono stati ospitati al Palazzo del Bo, la sede storica dell'Università degli studi di Padova, nell'aula magna Galileo Galilei. I secondi due giorni, i laboratori, si sono tenuti presso la Facoltà di Scienze Statistiche, nel campus di Santa Caterina, un ex-convento del XIV secolo.

Il programma dei lavori è stato arricchito con una serie di eventi sociali che hanno favorito la conoscenza reciproca e lo scambio di idee in modo informale tra i partecipanti a CLEF 2010. In particolare la cena sociale è stata anche l'occasione per consegnare alla dr. Carol

<sup>12</sup> *Multilingual and Multimodal Information Access Evaluation: Proceedings of the International Conference of the Cross-Language Evaluation Forum (CLEF 2010)*, edited by Maristella Agosti, Nicola Ferro, Carol Peters, Maarten de Rijke, Alan Smeaton, Heidelberg: Springer 2010.

<sup>13</sup> *CLEF 2010 Labs and Workshops, Notebook Papers*, edited by Martin Braschler, Donna Harman, Emanuele Pianta, Trento: MINT srl, 2010, <http://www.clef2010.org/index.php?page=pages/proceedings.php>.

<sup>14</sup> TEL è un portale progettato per far interoperare le biblioteche nazionali europee; ad oggi sono 48 le biblioteche nazionali accessibili in linea attraverso TEL. <http://www.theeuropeanlibrary.org/>.

Peters il premio Tony Kent Strix Award<sup>15</sup> che ha vinto nel 2009 per il suo lavoro di ideatore e coordinatore del Cross-Language Evaluation Forum negli ultimi dieci anni.

### CLEF 2011 e prospettive future

CLEF 2011 seguirà un formato simile a quello introdotto con CLEF 2010 è cioè una conferenza seguita da laboratori e workshop.

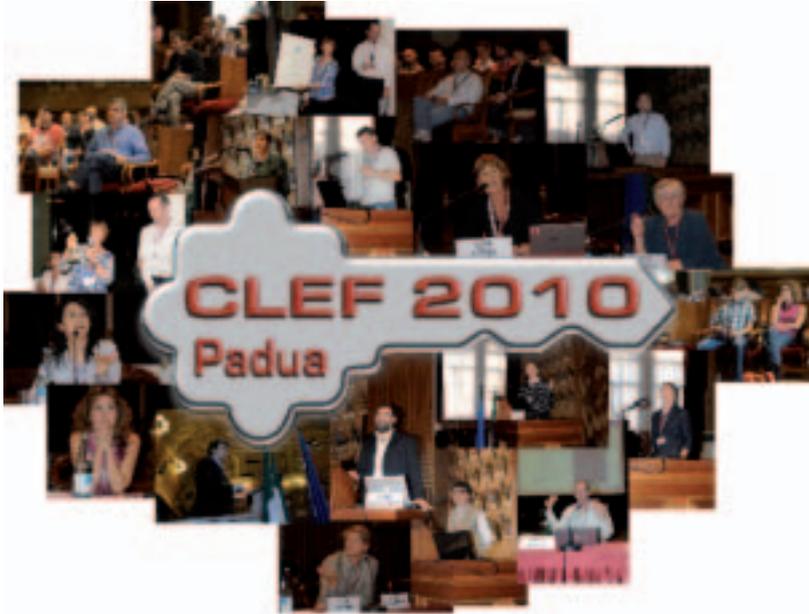


Figura 1. Alcuni momenti di CLEF 2010

CLEF 2010 è stato parzialmente supportato dal progetto PROMISE<sup>16</sup>, rete di eccellenza finanziata nel settimo programma quadro della Commissione Europea. Inoltre, per la prima volta, ha ricevuto il supporto finanziario di diverse organizzazioni e istituzioni: Google, Svizzera; Xerox Research Centre Europe (XRCE), Francia; Information Retrieval Facility (IRF), Austria; CELI Language Information Technology, Italia; Dipartimento di ingegneria dell'informazione dell'Università degli studi di Padova, Italia; Consorzio per la formazione e la ricerca in ingegneria dell'informazione in Padova (COFRIDIP), Italia.

CLEF 2011 verrà organizzato dall'Università di Amsterdam, in Olanda, nel mese di Settembre 2011.

Per CLEF 2012 e successivi, si progetta di istituire una procedura di selezione in modo tale che le organizzazioni e le istituzioni interessate possano candidarsi ad organizzare l'evento e competano per ospitarlo.

#### Ringraziamenti

*Il lavoro riportato in questo articolo è stato parzialmente supportato dalla rete di eccellenza PROMISE (contratto n. 258191), parte del settimo programma quadro della Commissione Europea.*

*The work reported has been partially supported by the PROMISE network of excellence (contract n. 258191) project, as part of the 7th Framework Program of the European Commission.*

<sup>15</sup> Tony Kent Strix Award, <http://www.ukeig.org.uk/awards/tony-kent-strix>.

<sup>16</sup> PROMISE, <http://www.promise-noe.eu/>

## Segnalazioni



# Antonella Agnoli, Le piazze del sapere

**Marta Cardillo**

ICCU

**D**opo i musei tocca alle biblioteche: già al termine degli anni Trenta il Guggenheim di Wright e successivamente, negli anni Settanta, il Pompidou di Piano e Rogers furono concepiti sulla base di una significativa svolta, in questo caso museologica. Qui spazi esclusivamente destinati alla tutela e all'esposizione delle opere si trasformarono in piazze coperte, luoghi d'incontro dove passeggiare, usufruire di servizi culturali e di ristoro; attività che si aggiunsero alle modalità tradizionali di fruizione dei beni culturali. Ora, sul versante bibliotecario, ecco chi ci ricorda che lo stesso compito può essere assunto anche dalle biblioteche pubbliche, anzi *deve* essere adottato per garantirne la sopravvivenza nel tempo. Antonella Agnoli, con *Le piazze del sapere*<sup>1</sup>, in libreria da circa un anno, ma ormai alla quarta ristampa vista la calda accoglienza dei lettori, parte dall'assunto che la "biblioteca pubblica", prevalentemente a carattere locale, debba perseguire quelle finalità per le quali è stata creata e che, con il tempo, si sono perse a causa del processo di irrigidimento che ha subito l'intera struttura. E lo fa filologicamente, rispolverandone la fisionomia ottocentesca, che consiste – come da manuale – in un «istituto della democrazia, centro fondamentale per lo sviluppo, educativo e sociale, degli individui e dei gruppi in og-

ni comunità locale» e che va a coprire «i settori sociali maggiormente a rischio di esclusione e di emarginazione e a estendere la cultura di servizio al di là delle proprie mura»<sup>2</sup>. Questo rimane l'unico modo, secondo l'autrice, per radicarsi in una società nella quale l'asse portante della cultura si è spostato dai contenuti alle relazioni, favorendo un impoverimento intellettuale di massa a cui l'istituzione pubblica deve, in qualche modo, rispondere. Di conseguenza anche le funzioni delle biblioteche debbono mutare da spazi per la lettura ad occasioni di incontro, di comunicazione, che rimangono l'unico strumento per poter avvicinare alla cultura chi altrimenti ne resterebbe escluso. Questa trasformazione deriva anche dal fatto che la società contemporanea fonda la propria economia su un modello non più basato principalmente sul processo di produzione, bensì sulla diffusione, nello spazio e nel tempo, della conoscenza, ossia sull'economia cognitiva: i servizi offerti e la loro interconnessione hanno un valore maggiore rispetto alla tradizionale fruizione biblioteconomica.

Inoltre, la nuova biblioteca pubblica si trova ad operare in un periodo in cui la funzione di quegli spazi urbani che prima erano adibiti alle relazioni sociali, *le piazze* per l'appunto<sup>3</sup>, è stata annullata, traslata a favore dei luoghi

<sup>1</sup> Antonella Agnoli, *Le piazze del sapere*, Bari-Roma: Laterza, 2009.

<sup>2</sup> *Biblioteconomia: guida classificata*, diretta da Mauro Guerrini, Milano: Editrice bibliografica, 2007, p. 726.

<sup>3</sup> «Le piazze sono il motore della democrazia» e «la biblioteca non può sfuggire alla crisi dei luoghi pubblici se non si dà un nuovo compito: trasformarsi in luogo di incontro, in un "piazza coperta" a disposizione dei grandi e dei piccoli, ricchi e poveri, zingari e cardinali» (Agnoli, *Le piazze* cit., p. 66, 67).

del profitto. Non è dunque azzardato che l'Agnoli, da un punto di vista metodologico, assimili la lezione di quei principi che già da più di un decennio hanno invaso lo spazio dei beni culturali e che fanno capo alla branca del marketing (disciplina alla quale ogni operatore culturale si è dovuto, volente o nolente, confrontare). Semplificando: per avere successo la strada più certa da percorrere è quella di osservare il consumatore per poi precederlo e guidarlo nelle scelte al fine di accontentarlo. E chi meglio dei supermercati ha saputo sfruttare questa metodologia? Scevra da pregiudizi culturali e seguendo una scia tracciata, nell'ultima decade, da una folta schiera di teorici<sup>4</sup>, Agnoli ritiene opportuno far riferimento a questa gamma di modelli presi direttamente da esempi "altri" rispetto a quelli tradizionalmente impiegati per le attività culturali. Perché sono luoghi socialmente frequentati. Perché assurgono dunque a modello di incontro e di reale confronto. E perché possono offrire spunti e suggerimenti a cui far riferimento per riuscire a decifrare e decodificare la struttura ed il linguaggio della società moderna alla quale le biblioteche si devono relazionare.

Contentezza, agio, simpatia, apertura: la nuova biblioteca pubblica, seguendo queste direttive, deve essere *consumer oriented* e *user friendly*. Tutto va potenziato verso questo scopo: architetture, arredamento, design, catalogo, orari, bibliotecari e personale. Una biblioteca «che abbia al centro le persone, non le collezioni»<sup>5</sup>, riporta la Agnoli in questo te-

sto che trasuda attenzione alla psicologia dei luoghi. Non più rigide sale di lettura ma spazi ibridi, accoglienti e plurimi, per una molteplicità di utenza e per una attenta ed amorevole cura di un pubblico variegato in età, sesso, abitudini, cultura e stato sociale. Una biblioteca sociale che esiste in uno spazio nuovo rispetto a quello che era prima, allineato a situazioni analoghe quali la *work-life balance*<sup>6</sup> o le ibridazioni commerciali che propongono servizi "spuri", come librerie che offrono ristorazione o ambienti per lo studio. Non esistono più barriere tra luogo e luogo, non più specificità in un divenire fluido di spazi delocalizzati. Il problema che ne consegue è dunque un problema collettivo della società postmoderna di cui la Agnoli è ben consapevole, «L'obiezione di molti colleghi è: "Che ne sarà della specificità della biblioteca? Noi non siamo un consultorio, né un ufficio postale, né una sala giochi, né un rifugio per i senza tetto"»<sup>7</sup>, ed al quale offre una risposta chiara e precisa: se le biblioteche non si adegueranno, diventando «anche tutto questo»<sup>8</sup>, moriranno perché, pragmaticamente, questa è la *realtà* della nostra società e non è possibile sopravvivere se non ci si rapporta ad essa. Ma è comunque possibile – si potrebbe obiettare – mantenere la propria specificità e nello stesso tempo offrire servizi che siano di forte impatto sociale e di elevata qualità. Questa situazione di rinnovamento che investe i luoghi deputati alla lettura invade anche i bibliotecari. D'altronde l'Agnoli non ha tutti i torti: la figura del bibliotecario, infatti, nell'immagina-

<sup>4</sup> Citiamo, tra gli altri, Rem Koolhaas. Uno tra gli architetti più produttivi dal punto di vista teorico, ci ha lasciato una sostanziale letteratura sul tema, dove il supermercato viene inteso come un luogo culturalmente sociale ed innovativo.

<sup>5</sup> Agnoli, *Le piazze* cit., p. XII.

<sup>6</sup> Un esempio per evidenziare come la proposta della Agnoli, in realtà, fa riferimento a sistemi comuni applicati in vari settori: l'erogazione di servizi utili a sostegno dei dipendenti sul posto di lavoro per una maggiore produttività implica la creazione di ambienti con «comfort e benessere ambientale, forme e colori rilassanti o energizzati a seconda dell'attività che si sta svolgendo» (Enzo Riboni, *Quando lavorare diventa un gioco: biliardo, playstation e calcio balilla entrano in ufficio*, «Corriere della Sera», 17 settembre 2010, p. 49).

<sup>7</sup> Agnoli, *Le piazze* cit., p. XII.

<sup>8</sup> *Ibidem*.

rio collettivo nazionale non spicca certo per entusiasmo<sup>9</sup>, mentre, per rispondere ad un'utenza che, inevitabilmente, oggi asserisce «Voglio tutto e subito»<sup>10</sup> e dove la figura professionale è scansata nella sua funzione di intermediario da una tecnologia utile ed intelligente, forse andrebbe rivista. A favore di gente motivata, intraprendente, anticonformista, creativa, audace e pronta a lavorare con differenti professionisti.

Per concludere il testo dell'Agnoli risulta un libro incisivo, anche se a momenti forzato<sup>11</sup> per la finalità che ha di voler intaccare una sclerotizzata e palese stanchezza culturale e un certo punto di vista che inevitabilmente pone a confronto la democratizzazione vs l'elitarismo. Anche il linguaggio utilizzato – sciolto, piacevole, talvolta divertente, anch'esso *user friendly* in una parola – è diretto a rafforzare l'obiettivo primario, che è quello

di intendere la biblioteca pubblica come lo strumento privilegiato per combattere i pregiudizi e rafforzare la visibilità urbana aumentando il “capitale sociale” del territorio. È sostenuto da dati analitici e da una serie di buone pratiche di varie biblioteche, tra cui ricordiamo, tra gli altri, gli Idea Store di Londra ed il San Giovanni di Pesaro, ai quali si potrebbe aggiungere la Mediatheque Villeurbanne. Realizzata da Mario Botta per il piccolo comune francese vicino Lione, infatti, è un ottimo esempio di servizi al pubblico ove l'attenzione, oltre ai libri e ai supporti mediali, si rivolge anche al settore artistico mediante il prestito di stampe d'autore contemporanee: è possibile, infatti, portarsi a casa un Takashi Murakami, un Daniel Buren oppure una fotografia di Gabriele Basilico da appendere in salotto e poi restituirla, dopo sei mesi, in mediateca.

<sup>9</sup> Cfr. il sito dell'Associazione Italiana Biblioteche (AIB) con la sezione *Librarians: Biblioteche e bibliotecari nella letteratura e nell'arte* (<http://www.aib.it/aib/clm/lett0.htm>).

<sup>10</sup> Agnoli, *Le piazze* cit., p. XII.

<sup>11</sup> È possibile trovare espressioni del tipo: «La biblioteca è un pezzo di città, dove per caso ci sono anche dei libri», Ivi, p. 158.

# Roberto Raieli, Nuovi metodi di gestione dei documenti multimediali\*

**Manuela Corbosiero**

Università degli studi di Roma "La Sapienza"

«Multimedia information retrieval is a cross-cutting field. Extending beyond the borders of culture, art, and science, the search for digital information is one of the major challenges of our time. Digital libraries, bio-computing & medical science, the Internet and social networking sites, streaming video, multimedia databases, cultural heritage collections and P2P networks have created a worldwide need for new paradigms and techniques on how to browse, search and summarize multimedia collections and more generally how to afford efficient multimedia content consumption»<sup>1</sup>.

Questa la premessa all'undicesima conferenza internazionale dedicata al *multimedia information retrieval* (MMIR), tenutasi a Filadelfia lo scorso marzo. L'appuntamento, sponsorizzato dall'ACM, Association computing machinery, è ormai istituzionalizzato da anni. Tuttavia le problematiche e le prospettive connesse ai sistemi di MMIR, discusse da ingegneri, informatici, matematici, esperti di *information technology* che operano in ambiti prossimi al *knowledge management*, risul-

tano ancora poco note tra coloro che gestiscono direttamente l'informazione nelle biblioteche, nelle mediateche, negli archivi o nei centri di documentazione.

Il lavoro di Roberto Raieli, esito di anni di studi e approfondimenti sul tema<sup>2</sup>, ha l'innegabile pregio di introdurre in maniera organica e sistematica la teoria e la tecnica del MMIR, cogliendone l'aspetto rivoluzionario in relazione a determinate prospettive sviluppate dalla biblioteconomia e dalla *library and information science*, e suggerendo una direzione di indagine concettuale e di riflessione che non può essere ignorata da bibliotecari e documentalisti. Anche perché il campo di interesse legato agli aspetti organizzativi e gestionali del MMIR è estremamente vasto e spazia dalla medicina alla musica, dall'ingegneria all'archeologia, dalla geografia alle arti e così via, ed è evidente che la questione della mediazione, disseminazione e fruizione di documenti multimediali acquisisce un'importanza strategica nell'ambito delle biblioteche ibride e digitali, laddove non vengono più gestiti solo documenti di tipo testuale, ma anche visivi, audiovisivi e sonori.

\* Roberto Raieli, *Nuovi metodi di gestione dei documenti multimediali: principi e pratica del multimedia information retrieval*, prefazione di Giovanni Solimine, Milano: Bibliografica, 2010.

<sup>1</sup> <http://riemann.ist.psu.edu/mir2010/index.html>.

<sup>2</sup> Tra i principali contributi citiamo *Multimedia information retrieval: metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione*, a cura di Roberto Raieli e Perla Innocenti, Roma: AIDA, 2004, *L'informazione multimediale dal presente al futuro: le prospettive del multimedia information retrieval*, a cura di Roberto Raieli, Roma: AIB Lazio, 2005 e un dottorato dedicato all'*Innovazione e sviluppo delle metodologie e delle tecniche della ricerca documentale*.

Il saggio, articolato in sette capitoli, è essenzialmente diviso in due parti. La prima, più ampia, esamina la teoria e le pratiche del MMIR. La seconda, più tecnica, ne presenta i sistemi e le concrete possibilità applicative. L'autore intende dunque in prima battuta contestualizzare le problematiche del MMIR in ambito nazionale e internazionale, considerando i rapporti con i principi della biblioteconomia e della documentazione, per segnare la possibile posizione teorica del nuovo sistema di trattamento dell'informazione e delineare una metodologia operativa. La discussione da un lato affronta i principi di indicizzazione e trattamento dell'informazione e dei documenti multimediali digitali, dall'altro tende a indicare e sviluppare quanto di innovativo c'è nell'odierna visione del trattamento dell'informazione. Se infatti l'*information retrieval* (IR) è tradizionalmente un sistema di ricerca terminologica di documenti di tipo testuale, applicato anche a documenti visivi, sonori, audiovisivi, il MMIR si caratterizza quale sistema organico di ricerca tramite testi, immagini e suoni, per documenti di tipo testuale, visivo, sonoro, audiovisivo.

L'aspetto più rivoluzionario e affascinante delle metodologie del MMIR consiste in un approccio alla ricerca di informazione, definito *content-based*, che punta direttamente al contenuto oggettivo dei documenti e che tende a superare i sistemi tradizionali di analisi e ricerca documentaria *term-based*, fondati sull'equivalente verbale per descrivere il contenuto concreto di un documento.

Se si volessero interrogare fonti di documenti testuali – fa notare l'autore – tramite mezzi di tipo non testuale, la confusione di linguaggi sarebbe considerata paradossale. Eppure nel caso in cui si interrogano tramite testo fonti documentarie di tipo visivo, sonoro o audiovisivo, lo scambio di linguaggi è ritenuto proficuo e op-

portuno. In realtà ogni genere di documento andrebbe archiviato e recuperato tramite gli elementi di linguaggio o metalinguaggio più adatti alla sua natura oggettiva e al tipo di informazione ricercata. Nei database in cui il contenuto dei documenti è principalmente testuale, è appropriato che le chiavi di accesso siano termini e frasi estratti dall'interno dello stesso contenuto. Nei database multimediali, invece, è semplificato e impreciso attribuire dall'esterno una descrizione testuale a contenuti consistenti in elementi visivi, sonori o audiovisivi. La metodologia *content-based* del MMIR consente un reperimento più appropriato dei documenti, tramite i mezzi del linguaggio del loro stesso contenuto informativo – il riconoscimento, la somiglianza, i rapporti di misure e valori – utilizzando quali chiavi di ricerca le strutture, per esempio, o le forme, le figure, i colori, le linee, i suoni, i movimenti.

Raieli, entusiasta sostenitore delle metodologie del MMIR, non manca tuttavia di coglierne le criticità e analizzare la problematicità di un metodo che presenti esclusivamente strumenti per operare sui contenuti concreti dei documenti e non tenga conto degli schemi intellettuali e quindi dell'interpretazione classificatoria. Se il sistema, infatti, si rivela valido nel caso di un approccio contenutistico e oggettivo al documento, mostra la sua limitatezza in relazione a un approccio intellettuale e interpretativo. Il miglior grado di precisione nel trattamento e allestimento di iter di recupero dei documenti multimediali, allora, si ottiene nella combinazione di tecniche e tecnologie di ricerca che integrano la rappresentazione del contenuto attraverso gli elementi multimediali con la definizione dei concetti attraverso termini semantici, in maniera da cogliere il senso dell'opera rappresentata nel documento nella sua totalità, attraverso la considerazione simultanea delle qualità sensibili e intellettuali,

di aspetto e significato, in ultima analisi concrete e astratte. Ed è questo ciò cui i più progrediti sistemi di MMIR devono tendere.

Punto di forza di *Nuovi metodi di gestione dei documenti multimediali* è sicuramente l'ampio respiro della trattazione che non si esaurisce in una descrizione tecnica, ma si sostanzia di un approfondito studio della letteratura professionale.

Raieli spazia con disinvoltura dalla citazione letteraria, all'analisi filosofica, attraverso una prospettiva epistemologica e un metodo interpretativo di chiara matrice er-

meneutica, con un'attenzione costante per le tematiche e le prospettive della biblioteconomia, assicurando al saggio quel carattere di universalità che ne fa uno scritto di prim'ordine.

Conclude il lavoro un'apprezzabile bibliografia, organizzata in sei sezioni tematiche. La prima è dedicata al MMIR nel complesso. Seguono gli aspetti specifici: *text retrieval*, *visual retrieval*, *video retrieval*, *audio retrieval*. Infine l'ampia serie di scritti dedicati ad argomenti collegati al MMIR.





### **Pubblicazioni recenti**

Futuro delle memorie digitali e patrimonio culturale, a cura di Vittoria Tola e Cecilia Castellani (2004)  
ISBN 978-88-7107-110-7

Memorie digitali: rischi ed emergenze, a cura di Alessandra Ruggiero e Vittoria Tola (2005)  
ISBN 978-88-7107-111-4

Linee guida per registrazioni d'autorità e di rinvio (2005)  
ISBN 978-88-7107-105-3

Linee guida per la digitalizzazione del materiale fotografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale fotografico (2005)  
ISBN 978-88-7107-112-1

EAD: Descrizione archivistica codificata: dizionario dei marcatori, a cura di Giovanni Michetti (2005)  
ISBN 978-88-7107-115-2

MAG: metadati amministrativi e gestionali: manuale utente, a cura di Elena Pierazzo (2006)  
ISBN 978-88-7107-117-6

Linee guida per la digitalizzazione del materiale cartografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale cartografico (2006)  
ISBN 978-88-7107-116-9

Inter omnes: contributo allo studio delle marche dei tipografi e degli editori italiani del XVI secolo (2006)  
ISBN 978-88-7107-114-5

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 5: D (2006)  
ISBN 978-88-7107-113-8

Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti, a cura del Gruppo di lavoro sulla digitalizzazione di bandi, manifesti e fogli volanti (2006)  
ISBN 978-88-7107-118-3

ISBD(CR): International Standard Bibliographic Description for Serials and Other Continuing Resources: revisione dell'ISBD(S) (2007)  
ISBN 978-88-7107-029-2

Miniature e decorazioni dei manoscritti, a cura di Emilia Ambra, Angela Dillon Bussi, Massimo Menna (2006)  
ISBN 978-88-7107-119-0

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 6: E-F (2007)  
ISBN 978-88-7107-125-1

Il libro italiano del XVI secolo: conferme e novità in EDIT16: atti della giornata di studio, a cura di Rosaria Maria Servello (2007)  
ISBN 978-88-7107-124-4

OAIS: Sistema operativo informatico per l'archiviazione, a cura di Giovanni Michetti (2007)  
ISBN 978-88-7107-126-8

ISBD(CR): International Standard Bibliographic Description for Serials and Other Continuing Resources: revisione dell'ISBD(S) (2007)  
ISBN 978-88-7107-104-6

Regole italiane di catalogazione: REICAT, a cura della Commissione permanente delle regole italiane di catalogazione (2009)  
ISBN 978-88-7107-127-5

Requisiti funzionali per i dati di autorità. Un modello concettuale, a cura di Glenn E. Patton. Roma, Iccu (2010)  
ISBN 978-88-7107-129-9

IFLA: International Standard Bibliographic Description (ISBD), Roma, ICCU (2010)  
ISBN 978-88-7107-128-2

Per l'acquisto delle edizioni ICCU rivolgersi a:

*ICCU Ufficio vendite*  
Viale Castro Pretorio, 105  
00185 Roma  
T +39 06 4989468  
F +39 06 4959302  
e-mail: [venditapubbl@iccu.sbn.it](mailto:venditapubbl@iccu.sbn.it)  
<http://iccu.sbn.it>

*Editrice Bibliografica*  
Via Bergonzoli, 1/5  
20127 Milano  
T +39 02 28315996  
F +39 02 28315906  
e-mail: [bibliografica@bibliografica.it](mailto:bibliografica@bibliografica.it)  
<http://www.bibliografica.it>



# ICCU

Istituto centrale per il catalogo unico  
delle biblioteche italiane e per le informazioni bibliografiche  
<http://www.iccu.sbn.it>

Copyright © ICCU - Roma

Finito di stampare nel mese di dicembre 2010